

# **Introduction To S-Plus 2000**

*By Fairouz Makhlouf*

Statistical Support Staff  
Center for Information Technology  
National Institutes of Health



# Contents:

• Getting Help	3
• Working with the Graphical Interface	41
• Data Import and Manipulation	66
• Graphics In S-PLUS	113
• Basic Statistical Models in S-PLUS	192
• Exporting Data	261
• History Log & Script Files	278
• Exercises	294
• Solution	305

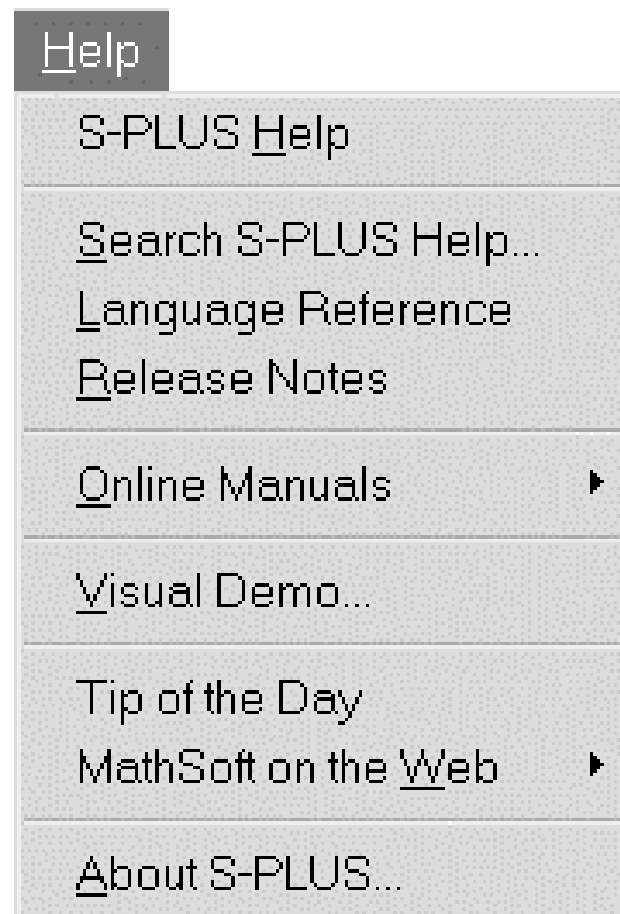
# Getting Help

# Contents:

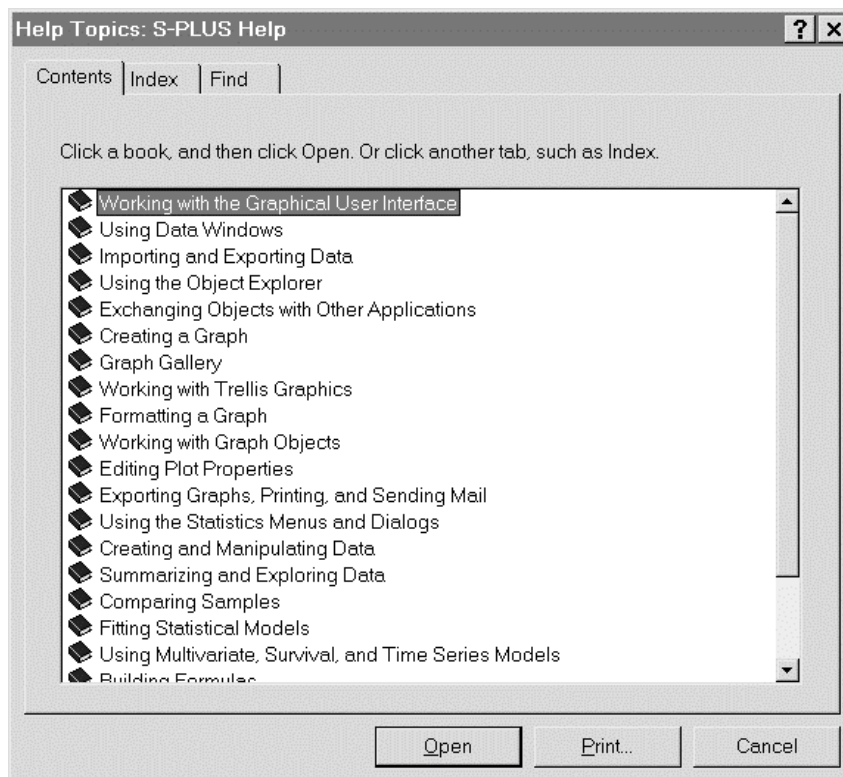
- How To Get Help???
- How To Get Support???
- What are the Learning Resources?
- What is New in S-plus 4.0?
- What is New in S-plus 4.5?
- What is New in S-plus 2000?

# How to get Help?

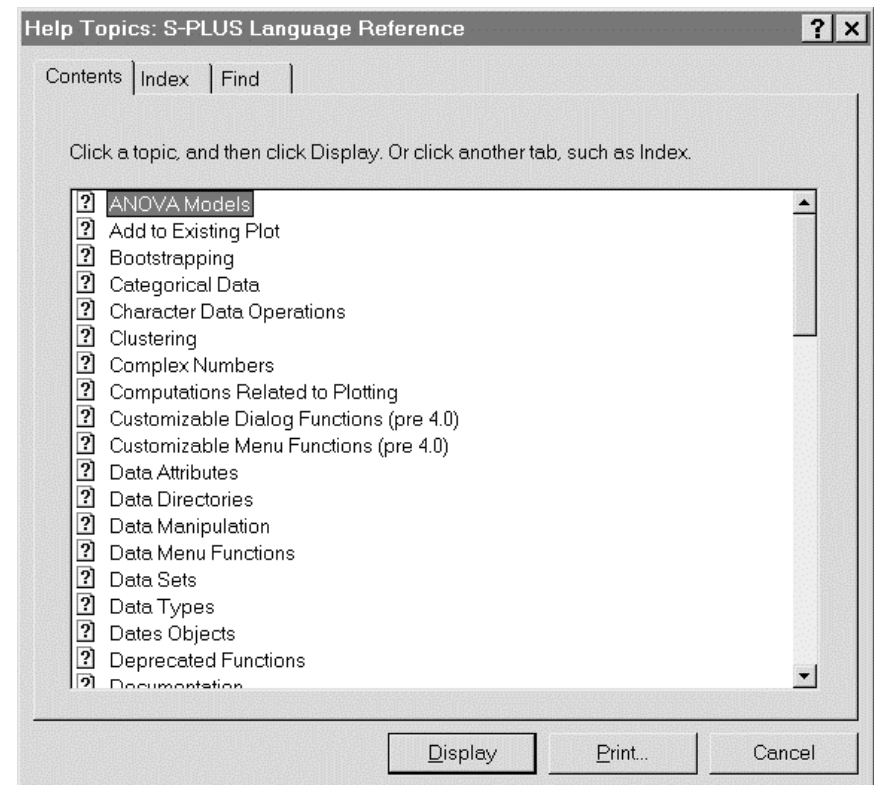
## ★ On-Line Help



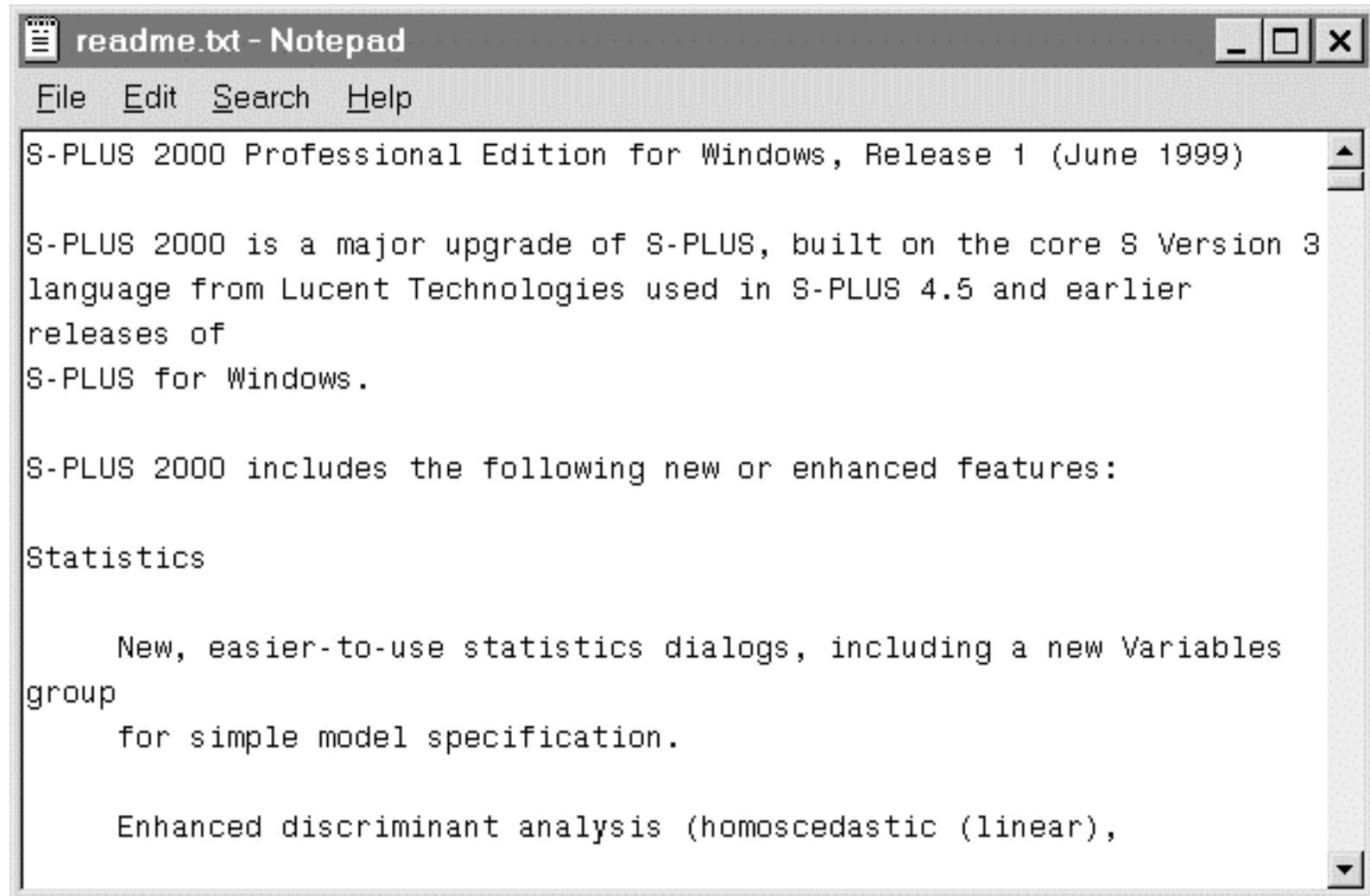
⊠ **S-PLUS Help:** How to use the graphical user interface



⊠ **Language Reference:** Details on each function on the S-Plus language



⊗ **Release Notes:** Details on the release of S-plus on your machine.



```
readme.txt - Notepad
File Edit Search Help
S-PLUS 2000 Professional Edition for Windows, Release 1 (June 1999)

S-PLUS 2000 is a major upgrade of S-PLUS, built on the core S Version 3
language from Lucent Technologies used in S-PLUS 4.5 and earlier
releases of
S-PLUS for Windows.

S-PLUS 2000 includes the following new or enhanced features:

Statistics

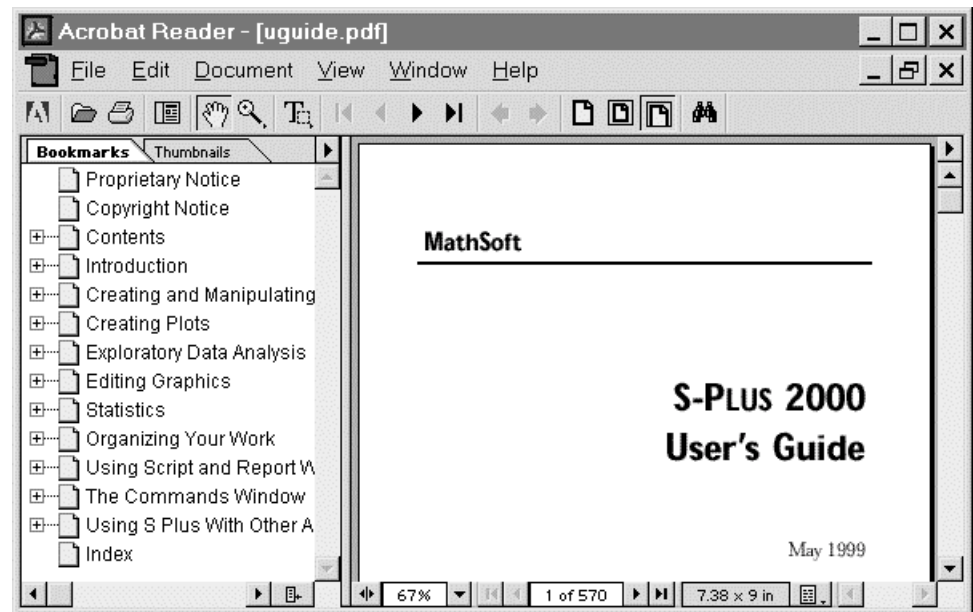
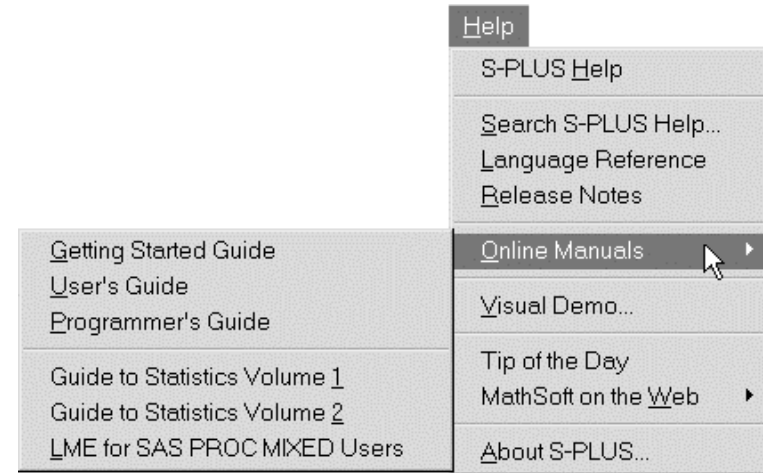
    New, easier-to-use statistics dialogs, including a new Variables
group
    for simple model specification.

    Enhanced discriminant analysis (homoscedastic (linear),
```

✉ **On-Line Manuals** : S-plus manuals are provided on line.

- To view the on-line manuals you need to install **Acrobat Reader**, which can be installed as an option during the installation process.

- It is useful to turn on bookmarks when using the **Acrobat Reader**.



☒ **Visual Demo** : Helps users to get an idea about the new features of S-plus.



# How to get Support ?

- Contact **Me** by

**Tel:** 4-6248 (GO-CIT)

**E-mail:** fhilal@exchange.nih.gov

- Contact the **Technical Support** at MATHSOFT

**Tel:** (206) 283-8802 ext. 235

**E-mail:** support@statsci.com

# What are the learning resources ?

## S-Press

- It is a free quarterly newsletter about S-plus.
- It is mailed to the primary users of S-Plus.
- It can be accessed with a web browser when you visit <http://www.mathsoft.com/splus/spress/>

## S-News

- It is an electronic mailing list by which S-Plus users can ask questions and share information with other users.
- To get on the list, send the message subscribe to  
S-news-request@utstat.Toronto.edu

# StatLib

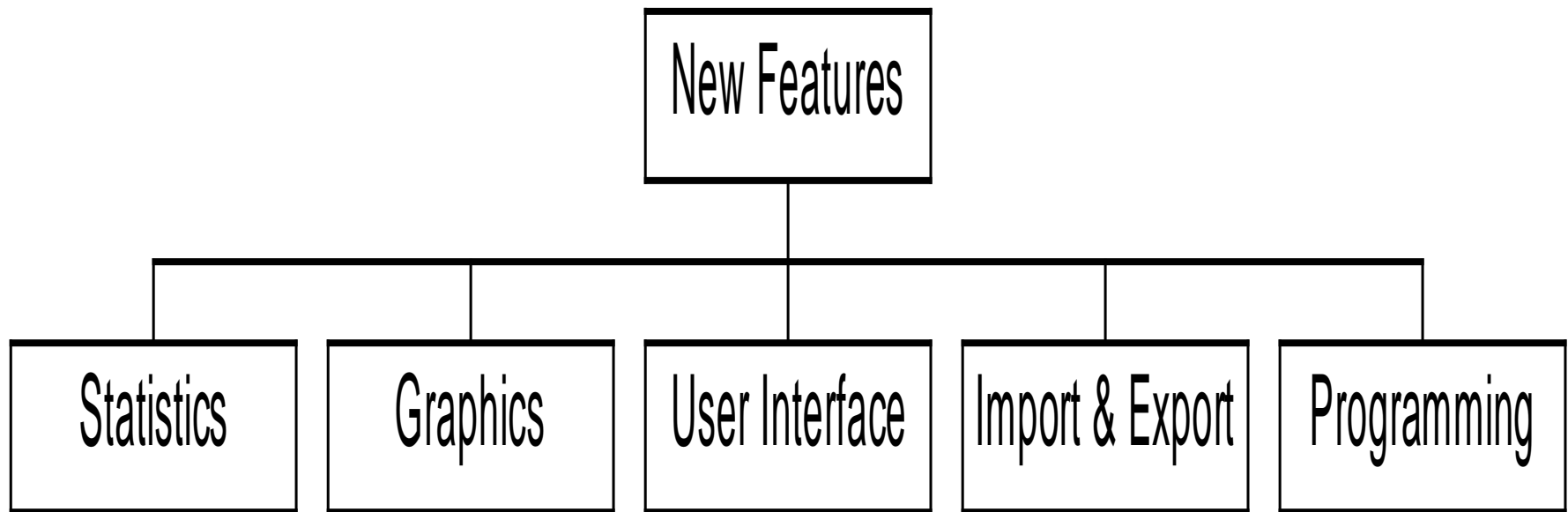
It is a system for distributing statistical software, data sets, and information by electronic mail, FTP and the World Wide Web.

- To access StatLib by FTP,
  - » open the connection to lib.stat.cmu.edu
  - » Login as a **anonymous**
  - » Send your e-mail address as the password
  - » The frequently asked questions are in /S/FAQ
- To access StatLib by e-mail
  - » send the message: **send index from S** to **statlib@lib.stat.cmu.edu**
  - » request any item in the StatLib with the request **send item from S** where **item** is the name of the item
- To access StatLib with a Web browser visit  
<http://lib.stat.cmu.edu/>

- **Books**

- **The Basic of S and S-Plus** (Statistics and Computing) ; Andreas Krause, Melvin Olson
- **Visualizing Data**; William S. Cleveland
- **Modern Applied Statistics With S-Plus** (Statistics and Computing); W. N. Venables, Brian D. Ripley.
- **Modern Applied Biostatistics Methods: Using S-Plus**; Steve Selvin.

# What Is New In S-PLUS 4.0?



# **Statistics:**

- Bootstrap and jackknife estimation
- Linear mixed models
- Nonlinear mixed effect
- Multiple comparisons
- Crisp and fuzzy clustering
- Monotheistic clustering
- Divisive and agglomerative methods

# **Graphics:**

- Multiple simultaneous 3D rotation views
- 2D projection in 3D space
- Interactive 3D view angle specification
- Drag-and-drop creation of Trellis graphics
- Point-and-click editable graphics
- Flexible page layout
- Multiple axis breaks

- Multiple-line text annotation
- Superscripts, subscripts, Greek letters and symbols
- Automatic creation of code to produce editable graphics
- Support for international ASCII characters

# **User Interface:**

- 32-bit application with full Windows 95, Windows NT and 3.1x running Win 32s compatibility
- Microsoft Office compatible interface
- Statistical menu and dialogs
- Editable graphics
- Customizable toolbars, menu and dialogs

- Tooltips, online help and tutorial
- Data windows for spreadsheet editing and display of data
- Multi-page Graph sheets
- Object Browser to organize data, graphs, functions and other objects
- Script editor for scripting and programming
- Report windows for easy handling of output

# **Import and Export:**

- Import SAS, SPSS, Excel and other file formats
- DDE, OLE client and server, graph documents
- Export graphs as .BMP, .EPS, .TIFF, .WMF, or other file formats
- Link data from Excel or other OLE spreadsheets

- Embed S-PLUS graphics in Microsoft Word and edit in place
- Create PowerPoint slides from S-plus graphics automatically

# **Programming:**

- Validation suite of tests cases to confirm accuracy of S-plus output
- OLE Automation client and server
- Edit and run scripts

# **What Is New In S-PLUS 4.5?**

# **Efficiency Improvements**

- Faster start-up
- Faster Data entry

# Statistics

- New robust regression method available via menu system and through lmRobMM function
- Power and sample size calculation
- Enhanced parametric survival (accelerated failure time) estimation.
- Type III sums of squares for ANOVA
- Bootstrap and jackknife estimation now available through the menu system.
- Clustering methods are now available through the menu system

# Graphics

- Interactively select and highlight points
- Redraw excluding selected points
- Linked Highlighting in scatter plot
- Interactively rescale axes
- Trellis drill-down: select one panel of a Trellis graphic and create a full size copy
- Color scale Legends

# User Interface

- Select Data Dialog at startup for easy data selection
- More data manipulation dialogs
  - Recode
  - Split
  - Stack
  - Subset
  - Transform
  - Transpose
  - Factorial Design
  - Orthogonal Array Design
  - Set Dimensions

- Improved Insert Graph dialog with enhanced thumbnails
- Excel add-in for S-Plus graphics

# **Import and Export**

- New Excel add-in to create S-Plus graphics from Excel
- Functions to generate HTML tables.

# Programming

- Dialogs now support ActiveX controls
- Enhanced automation support, including many new included examples
- Enhanced editing features in script windows includes automatic delimiter matching and auto-indent feature

# **License Management**

- License manager for network version

# **What Is New In S-PLUS 2000 ?**

# Statistics

- New easier-to-use statistics dialogs, including a new **Variables** group for simple model specification.
- Enhanced discriminant analysis.
- Enhanced LME/NLME.
- Enhanced Survival Analysis.
- Enhanced Quality Control charts.

# Data

- Import data from financial databases: Bloomberg, FAME and LIM.
- Data manipulation capabilities and dialogs have been greatly enhanced.
- DataTips now show column type and column description.
- New Data.sheet data object has been added.
- Save, load, and create data objects via the **File** menu.

# New Object Explorer

- View an outline of your data objects and your documents in the left pane. Expand and contract the outline as you do in the Windows Explorer.
- View the contents of any object in the right pane, along with explanatory information.
- Add your own folders to the **Object Explorer**. Add objects to it by filtering or by drag-and-drop.
- Select data columns from the right pane of the Object Explorer and graph them with the plot Palettes.
- View and manipulate time series objects.

# Graphics

- DataTips and labeling have been added to scatter plots.
- New plot Types have been added to the 2D palette.
- New **Default 2D Axis** toolbar option and new **Insert Graph** dialog let you specify 2D axis types before you create your plot.
- Embed data into **Graph Sheet** so that it can be saved as a single file.
- Extract data from **Graph Sheet** to link the plots to data.

# **More New Stuff**

- **Saving and Loading the workspace**
  - Workspace files can be saved and loaded. Each workspace file contains a complete description of the workspace, including sizes and positions of open documents and user-specified path.
- **Enhanced History Log**
  - By default, the history log is now considered, showing quick summaries of your actions. You can choose to see full command, if you prefer.

# More New Stuff

- **Automation**
  - Greatly enhanced automation server capabilities, especially for using C++ and Java to write automation clients for S-PLUS.
  - Enhanced automation client capabilities.
  - New automation help capabilities.
- **GUI customization and deployment**
  - Drill down capability in **Object Explorer** for **FunctionInfo** objects and for properties makes dialog creation much easier.

# **More New Stuff**

- **Interacting with other products**
  - New S-Plus component for Mathcad.
- **Improved Speed**
  - Start-up speed up 50% faster (depending on the machine)
- **Y2K compliance**
  - S-Plus 2000 for windows/NT has had complete Y2K test coverage.

# **More New Stuff**

- **User contributed libraries**
  - **from Brian Ripley and William Venables, and from Frank Harrell**
- **Documentation**
  - The documentation set has been revised and expanded.
  - Tips of the Day to improve your efficiency have been added.
  - New sample files have been added to the samples directory:
    - **Many sample Graphs and data sets**
    - **New automation examples**

# **Working With The Graphical User Interface**

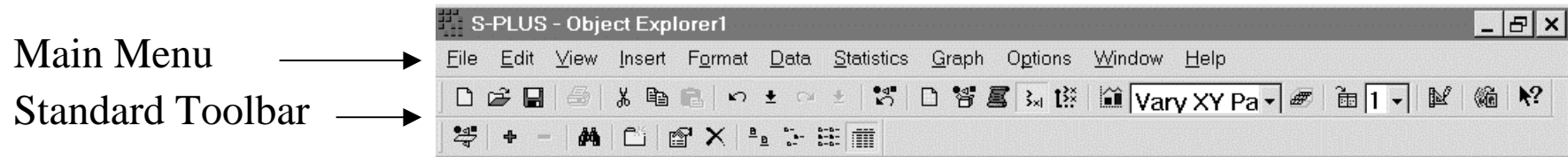
# Objectives:

- To get familiar with the following
  - S-plus Windows
  - S-Plus Menus
  - S-Plus Dialogs
  - S-Plus Palettes and Toolbars

# **S-PLUS Windows**

- Object Browser
- Data Window
- Graph Sheet
- Command Window
- History Log

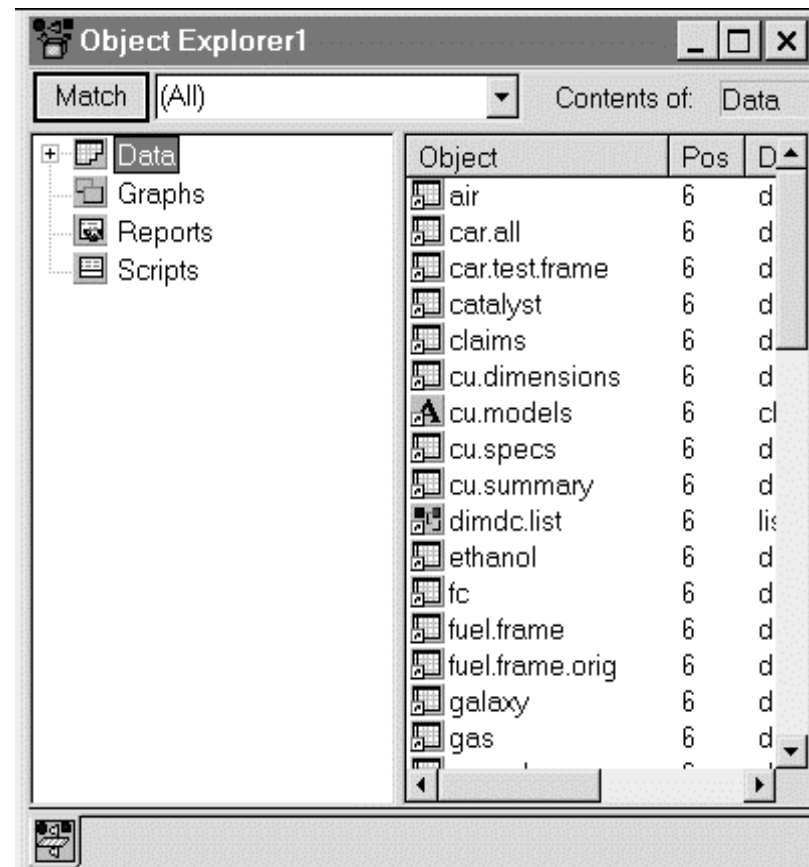
# Main Menu & Standard Toolbar



- **Main Menu:** Menu items are dependent on the type of window you are working on.
- **Standard Toolbar:** Toolbars contain buttons that are shortcuts to menu selection

# Object Browser

- It provides a detailed map of S-Plus.
- It displays the data sets, graphs, functions, and other objects in the S-plus session.



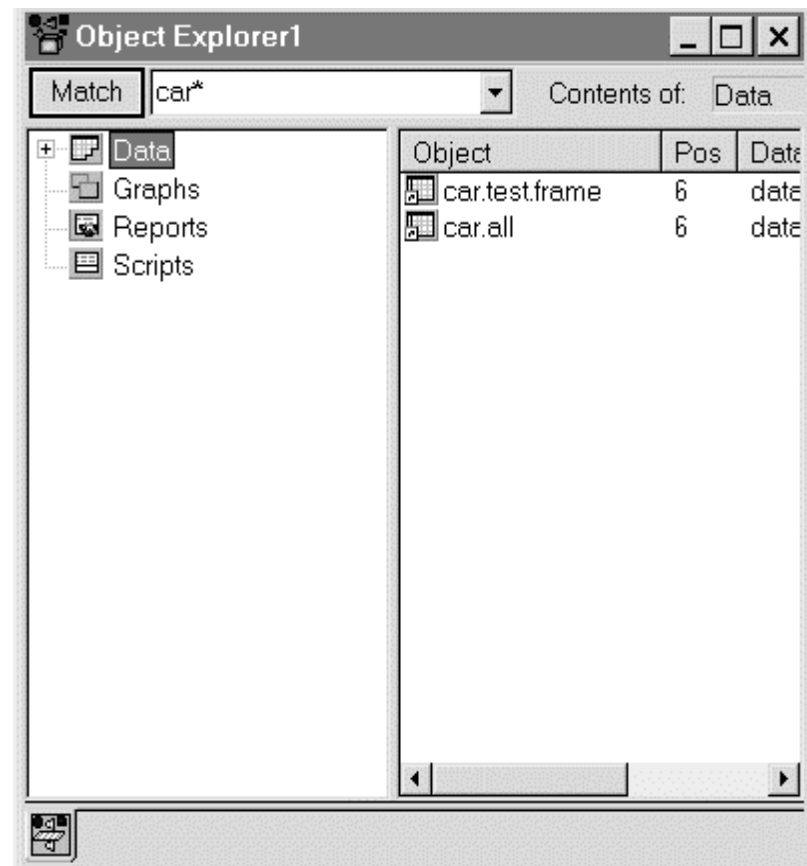
# Finding Objects Using the Object Browser

- **Two Ways**

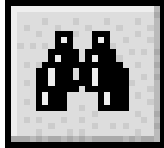
- Using the “Match” option to find objects in the current database.
- Using the ‘Find Objects’ option to find objects across multiple databases and directories.

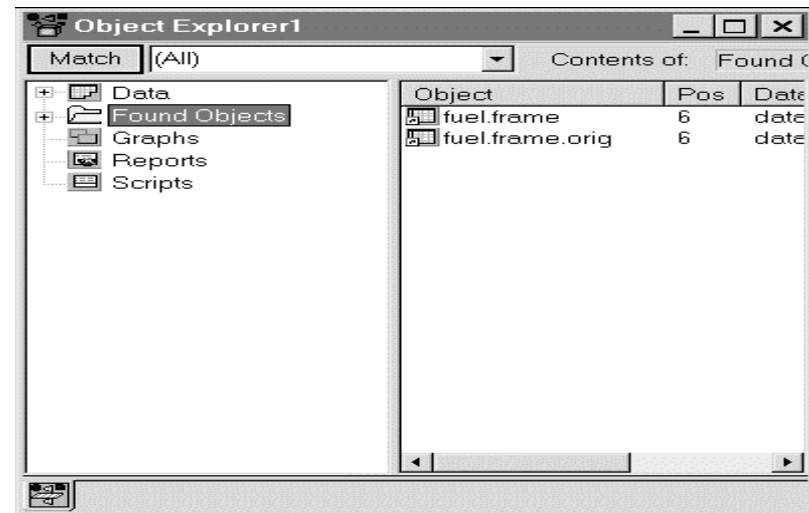
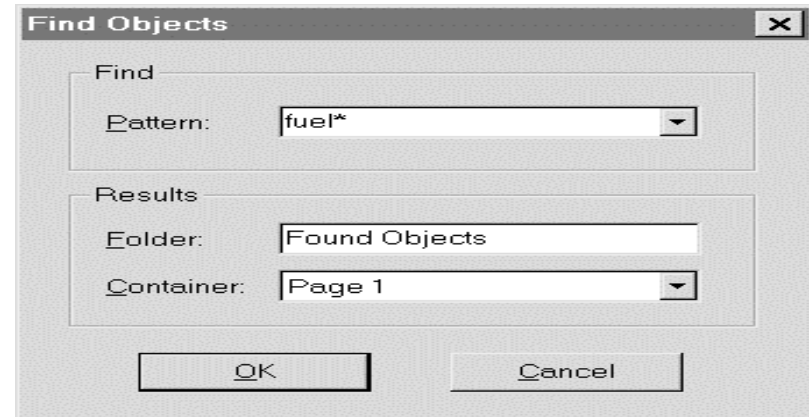
# Example: Using the “Match” option

- Find the objects that starts with “**car**”
- In the Match field, type the name “car\*”



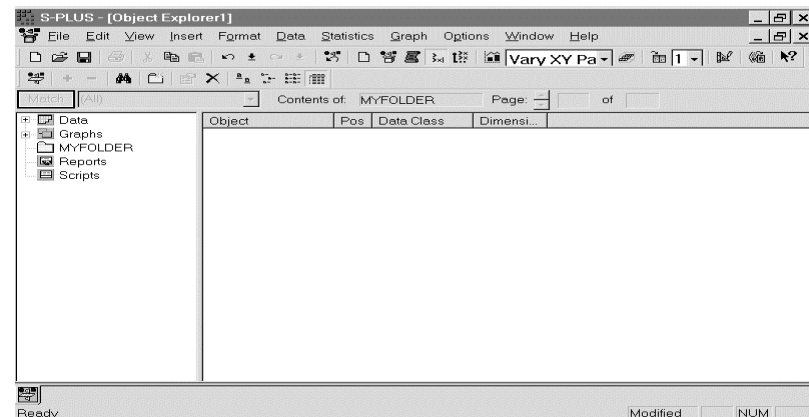
# Example: Using the “Find” option

- Find the objects “fuel”
  - Click on the “**Find Objects**” icon on the Object Browser toolbar 
  - Type “fuel\*” in the Pattern field
  - The “**Find Objects**” option locates “fuel” and locate it in the “**Found Objects**” folder.



# Creating a Folder in the Object Browser Page

- Right-click in the left pane of the OBP
- Select “**Insert Folder**” from
- A folder is inserted
- Enter a folder name

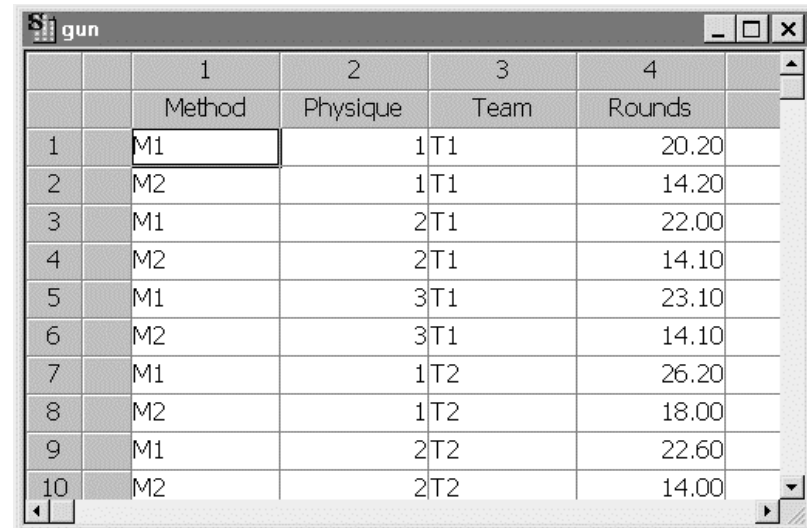


# Adding an Object to a Folder

- Two ways
  - Drag the icon of an object onto the folder
  - Use copy and Paste
- Show Examples in Class

# Data Window

- It displays data sets in an editable spreadsheet format.
- It makes it easy to create a graph, explore data and perform statistical analysis and modeling.



The screenshot shows a window titled 'gun' containing a spreadsheet. The spreadsheet has 10 rows and 4 columns. The columns are labeled 'Method', 'Physique', 'Team', and 'Rounds'. The rows contain data for 10 different entries, numbered 1 to 10 in the first column. The 'Method' column contains 'M1' or 'M2'. The 'Physique' column contains '1T1', '2T1', '3T1', '1T2', or '2T2'. The 'Team' column contains '1T1', '2T1', '3T1', '1T2', or '2T2'. The 'Rounds' column contains numerical values: 20.20, 14.20, 22.00, 14.10, 23.10, 14.10, 26.20, 18.00, 22.60, and 14.00.

	1	2	3	4
	Method	Physique	Team	Rounds
1	M1		1T1	20.20
2	M2		1T1	14.20
3	M1		2T1	22.00
4	M2		2T1	14.10
5	M1		3T1	23.10
6	M2		3T1	14.10
7	M1		1T2	26.20
8	M2		1T2	18.00
9	M1		2T2	22.60
10	M2		2T2	14.00

**Note:** To select more than one column, select the first column and then CTRL-click on the other one. You can select as many column as you want.

# Hot Spots in the Data Window

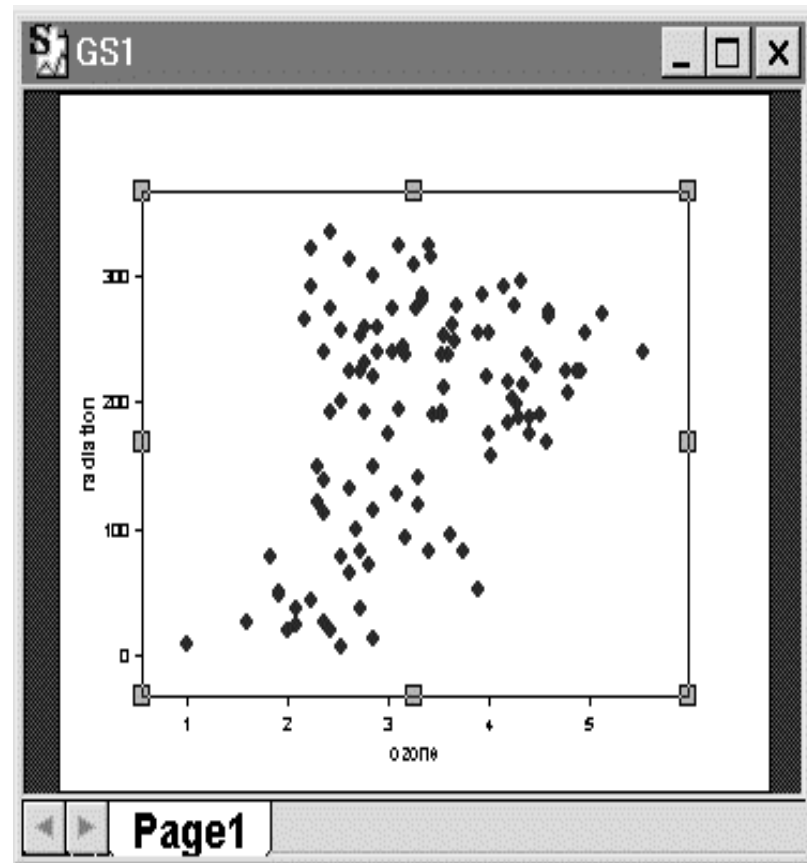
- ✦ To select the entire data frame
- ⬇ To select a column
- ↶ To move a column
  - Select the column
  - Position your pointer in the column and drag

# Data Tips

- Shows the column Type and column description.
  - Point the mouse cursor at the column number to see the column type.
  - Point the mouse cursor at the column name to see the column description.

# Graph Sheet

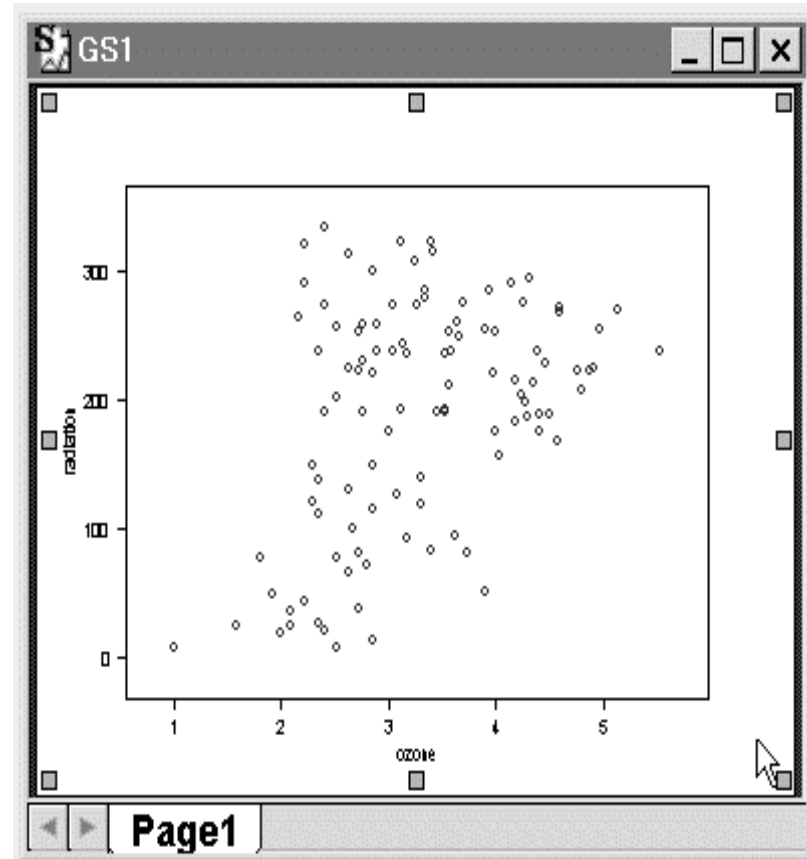
- Every plot in S-Plus is held in a Graph sheet.
- A graph sheet can contain one or more graphs.
- Graphs are object oriented which allows us to edit them.



- Specific areas of the graph sheet are identified by different names.
  - Graph Sheet: The actual page the plot is drawn on.
  - Graph Area: The area outside the plot axis but within the Graph sheet.
  - Plot Area: The area of the plot including the axis.
  - Plot: The actual points that form a plot.

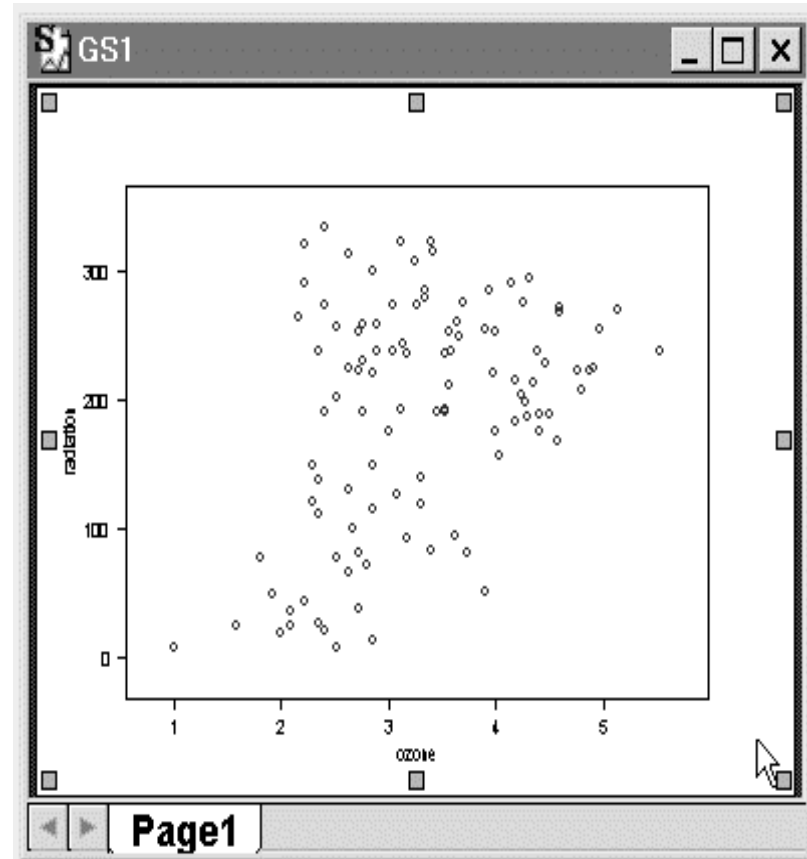
# Hot Spots in the Graph Sheet

- To select the graph sheet you need to position your pointer near the edge of the sheet and click. A green knobs will appear around the Graph sheet to identify that it has been selected.



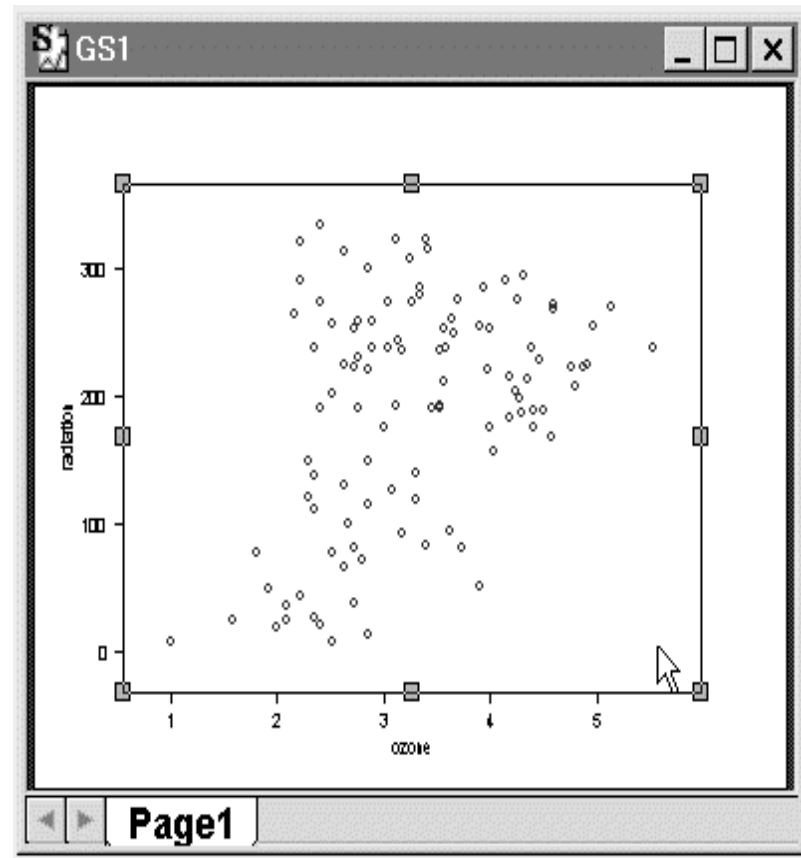
# Hot Spots in the Graph Sheet

- To select the graph sheet you need to position your pointer near the edge of the sheet and click. A green knobs will appear around the Graph sheet to identify that it has been selected.



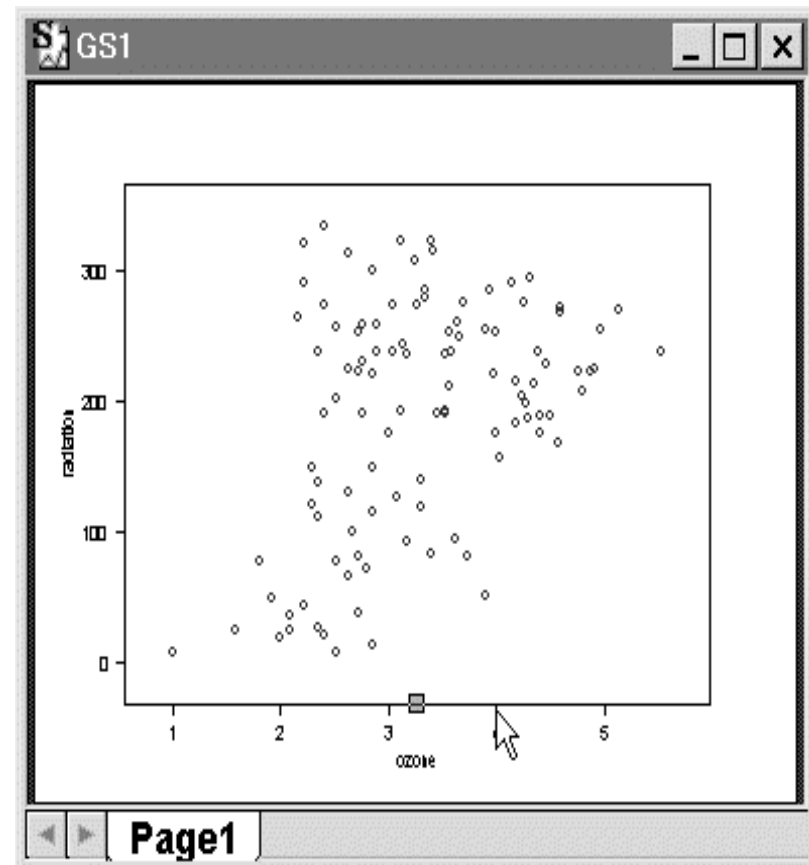
# Hot Spots in the Graph Sheet

- To select the plot area you need to position your pointer in the right-bottom corner inside the plot area and click. A green knobs will appear around the Graph area to identify that it has been selected.



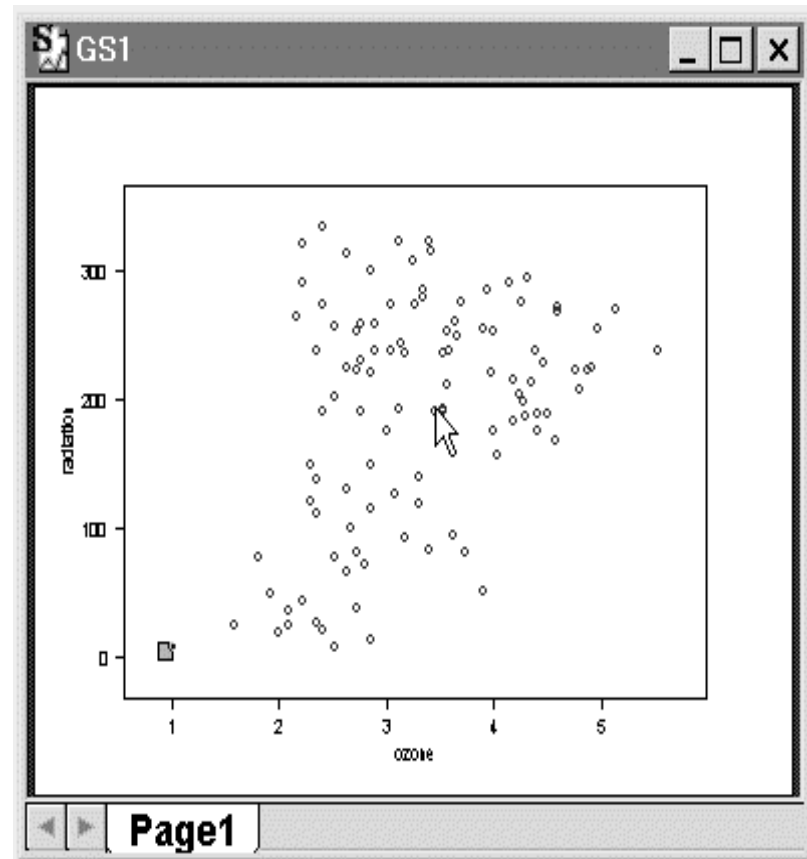
# Hot Spots in the Graph Sheet

- To select any of the axis you need to position your pointer directly on a tick mark and click. A single green knob will appear at the center of the axis to identify that it has been selected.



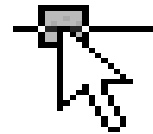
# Hot Spots in the Graph Sheet

- To select a plot you need to position your pointer directly on a data point and click. A single green knob will appear near the bottom of the plot to identify that it has been selected.



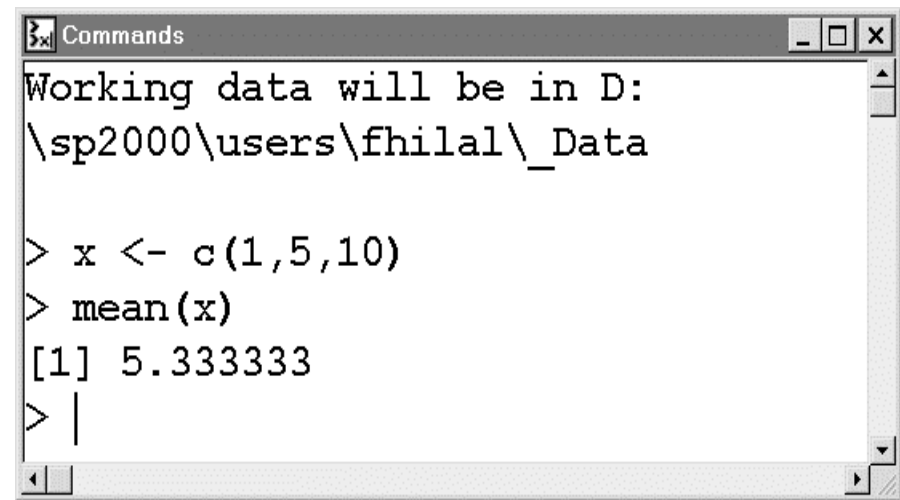
# Hot Spots in the Graph Sheet

- To open a dialog corresponding to the selected S-Plus object, double click on the green knob.



# Command Window

- It allows the users to access the S-Plus programming language to create new functions or to modify existing ones.




```
Commands
Working data will be in D:
\sp2000\users\fhilal\_Data

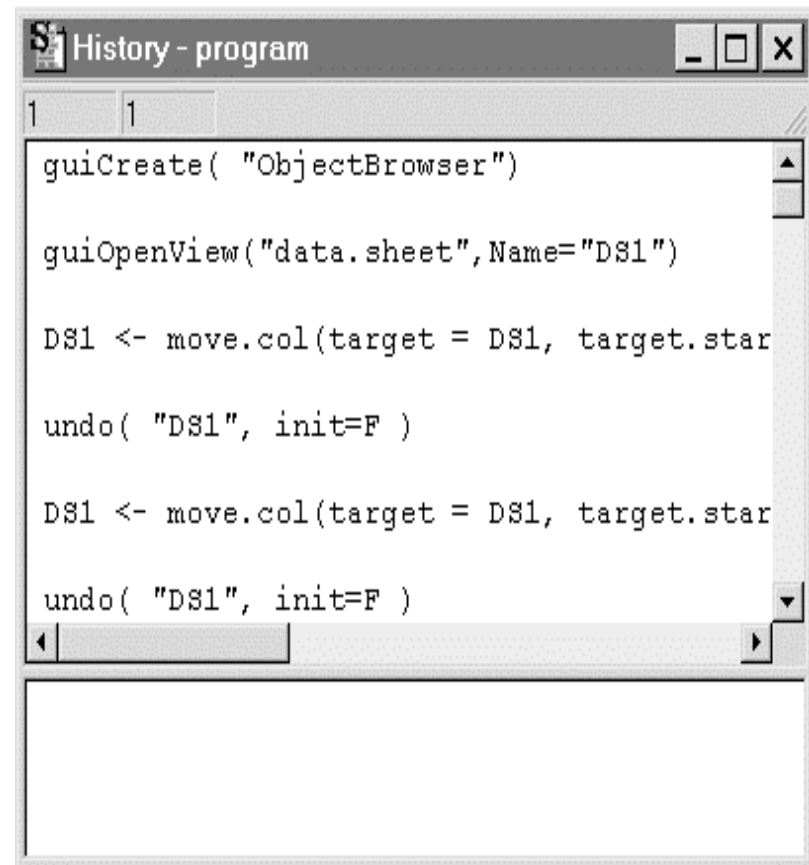
> x <- c(1,5,10)
> mean(x)
[1] 5.333333
> |
```

# History Log

- **Notes**
  - Everything you do in S-Plus is recorded behind the scenes in the S programming Languages
  - The generated code can be viewed automatically by the History Log
  - The History Log can be saved, edited and re-executed to perform repetitive tasks

# Display Commands from the current session

- Click on the History Log button 
- Close the “**History Log**” window
- Say No when you prompted to save the “**History Log**”



```
guiCreate( "ObjectBrowser")

guiOpenView("data.sheet",Name="DS1")

DS1 <- move.col(target = DS1, target.star

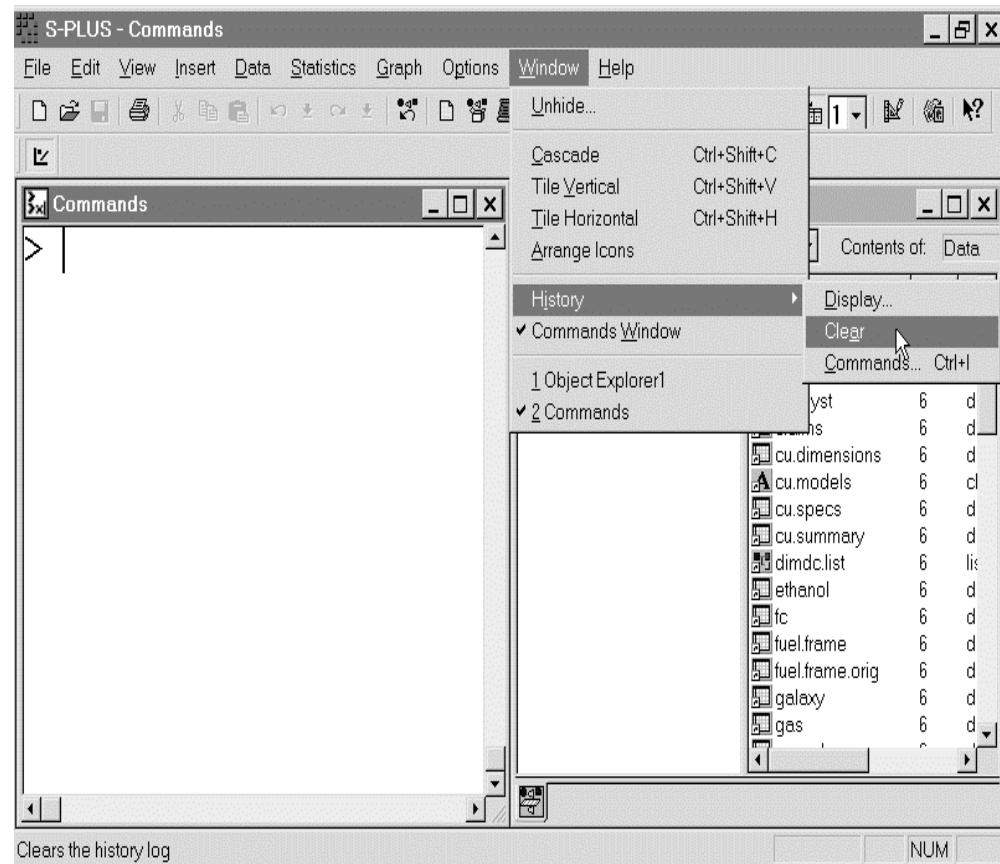
undo( "DS1", init=F )

DS1 <- move.col(target = DS1, target.star

undo( "DS1", init=F )
```

# Clear the existing “History Log”

- Click “**Window**”
- Click “**History**”
- Click “**Clear**”



# **DATA**

## **Import & Manipulation**

# Contents:

- Data Modes
- Data Types
- Importing Data
- Editing Data
- Handling Missing Data
- Creating New Data
- Adding Data to Existing Data Frames
- Merging Data Frames

# Data Modes

**Some basic modes of data in S-Plus are :**

- Numeric
- logical
- character

As well as

- Missing values
- Null values

# Numeric:

- It is used to describe numbers.
- Algebraic computations , trigonometric transformations, etc. can be conducted with numeric variables.

# Logical:

- Logical values in S-PLUS are **T** (for true) and **F** (for false). There are no quotes around these values.
- You can create logical variables
  - By typing the T's and F's in the cells of a data window
  - By Using S-Plus logical operator to create new variables
- Logical operators in S-Plus
  - `==` equals, `!=` not equal
  - `<` Less than, `<=` Less than or equal
  - `>` Greater than, `>=` Greater than or equal
  - `&` Elementwise And
  - `|` Elementwise Or
  - `&&` Control And
  - `||` Control Or
  - `!` Unary Not

# Complex:

- They are specified in the form  $a+bi$  where  $a$  is the real part and  $b$  is the imaginary part.
- Examples:
  - $1+0i$
  - $3-2i$
  - $5.7+2.3i$

# Character:

- It is letters of the alphabet plus other symbols and numbers
- It is used
  - for labeling observations in graphs
  - subsetting data for models.
  - Function arguments may require character strings
- If numbers are present, they will act like character strings in that you cannot perform numerical computations using these cells.
- If you enter character strings in the data window, there is no need to use quotes. On the other hand, if you enter character strings using S-Plus expression, you must use quotes around each character string.

- You could use either the double-quote (“) or single-quote to create character strings. Make sure that they are balanced.

# Missing Values:

- The missing values in S-Plus is **NA**. There are no quotes around the **NA**.
- When you import data, the missing value symbol of the file you are importing will be automatically converted to **NA**.
- Empty cells are not allowed in S-Plus. You must use **NA** to fill that cell.

## **Empty Objects - Null Values:**

- It is a place-holder for an empty object. This object exist in S-Plus, but it contains no data.
- These objects could have a pre-specified mode.

# Data Types

- There are seven basic types of data objects in S-Plus that can be classified into two categories
  - Atomic objects that can contain values of only one kind. They are
    - Vector data objects
    - Matrix data objects
    - Array data objects
    - Factor data objects
    - Time Series data objects
  - Non-Atomic objects that contain values of all kinds. They are
    - List data objects
    - Data Frame Objects

# Vectors Data Objects:

- It is a set of numbers, character values, logical values, etc..
- Vectors must be of a *single mode*.

# Matrix Data Objects

- It is a two-way array. They are used to arrange values by rows and columns in a rectangular table.
- All of the values should be of a *single mode*.

# Array Data Objects

- They are a generalization of matrices. Its dimension can be extended to more than two dimensions.
- Arrays have four attributes: length, mode, dim, dimnames.

# Factors and Ordered Factors

- The Factor data type is used to represent categorical data.
- Some examples of Categorical data are gender and marital status.
- Factor data objects must be of a single mode.

# Time Series Data Objects

- It is a collection of observations made sequentially in time.
- There are three types
  - regularly spaced time series
  - calendar time series
  - irregularly spaced time series

# **Lists data Objects**

- It is an ordered collection of components.
- Each components can be any data object.
- Different list component can be of different modes.

# Data Frames Data Objects

- They are data objects that allows you to group data by variables (columns) regardless of their type. Note that all of the variables must be of the same length.

# Data Import

## Types of data files that S-PLUS can import :

- ASCII
- Formatted ASCII (FASCII)
- Microsoft Excel (versions 2.1 through Excel 97 .XLS files)
- Quattro Pro (.WQ1)
- Paradox Databases (.DB)
- Lotus Worksheets (.WKS .WK1, WK3, .WK4 AND .WRK)
- DBASE FILES (.DBF II, II+, III, IV files)
- FoxPro files (use the same import filter as dBase files above)
- Systat files (double or single precision .SYS files)

- SigmaPlot files (.JNB)
- SPSS files (.SAV)
- SPSS Export files (.POR)
- SAS files (.SD2)
- SAS Transport files (version 6.x. TPT). Some special export options may need to be specified in your SAS program. Using SAS Xport engine is recommended to read and write these files.
- Microsoft Access files (.MDB)
- Matlab (.MAT)
- S-PLUS Transport files (.SDD)
- STATA files (.FTA Versions 2.0 and higher)
- Gauss (.DAT files - automatically reads the related DHT file)

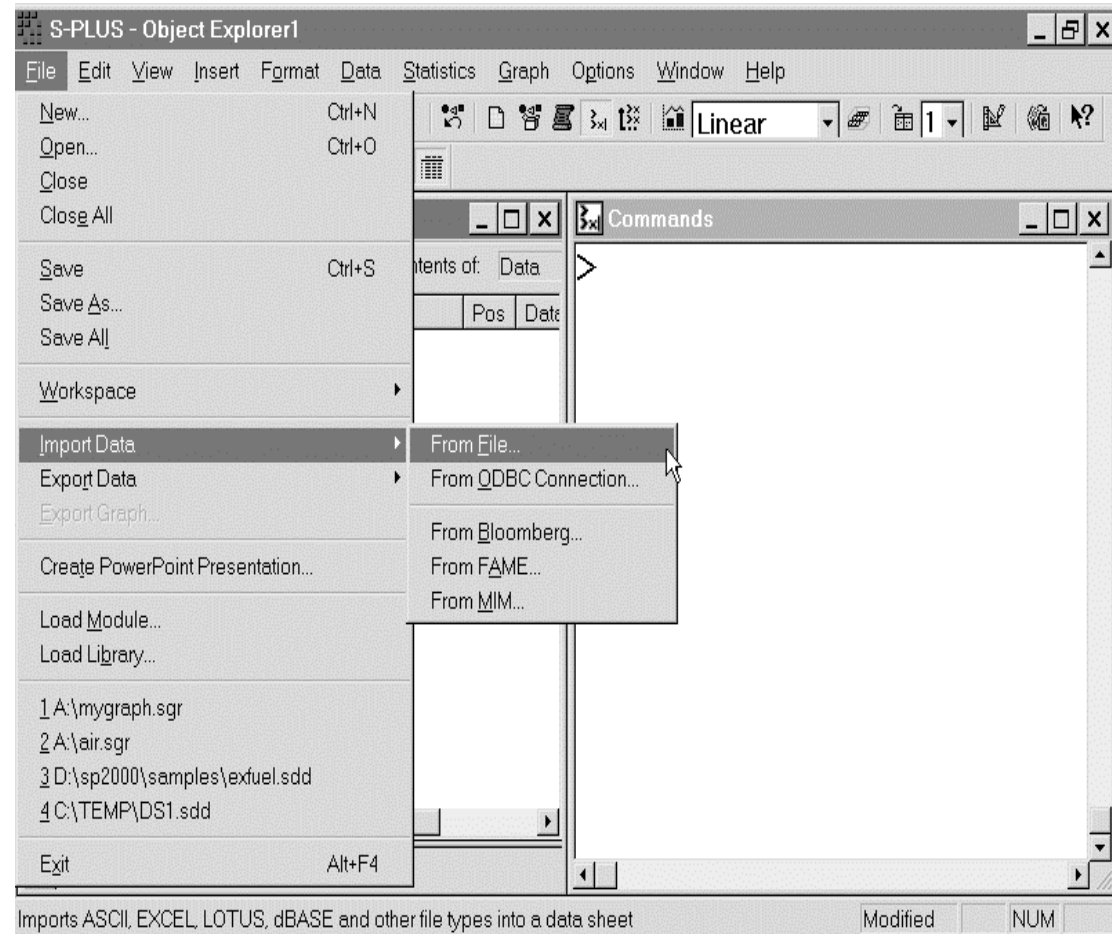
## **To import a data file we need to do the following:**

- From the file menu, choose Import Data
- Then From File
- In the File name box, type or select the name of the file you want to import.
- In the Files of type box, select the type of the file to import.
- In the Import To Box, specify the target data frame, starting column.
- Click Open to import the file.

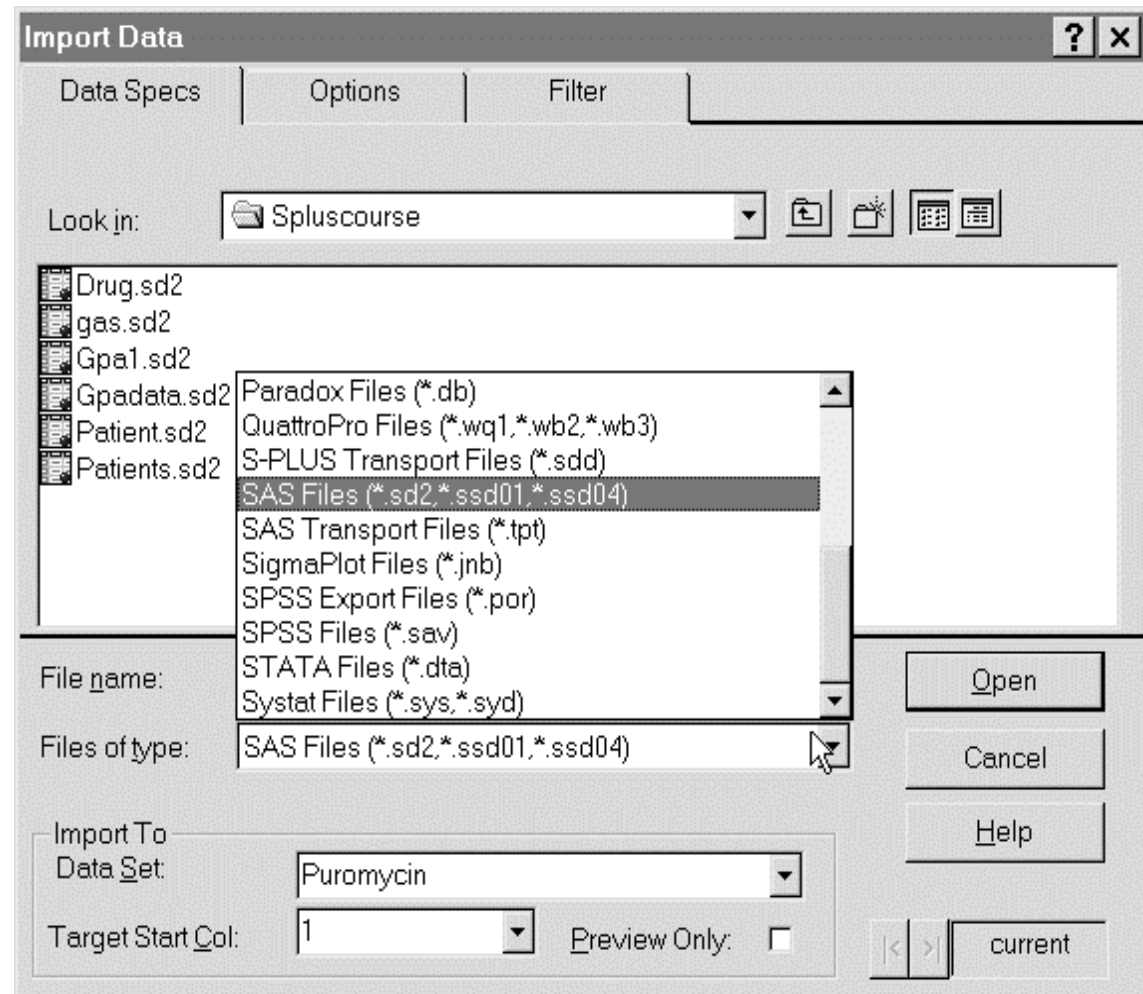
# Example: Importing DATA

Import Drug.sd2 data set into SPLUS.

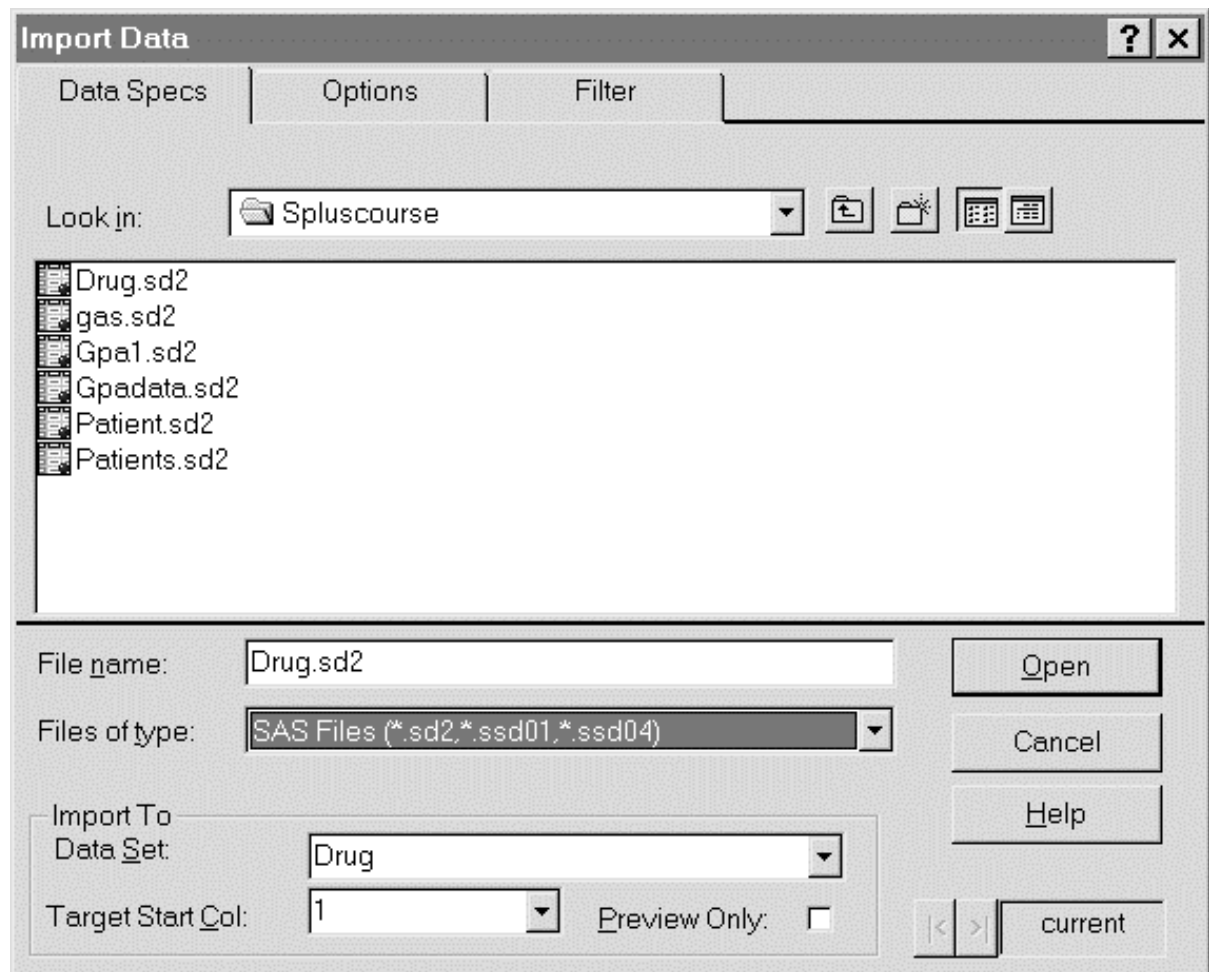
Step 1



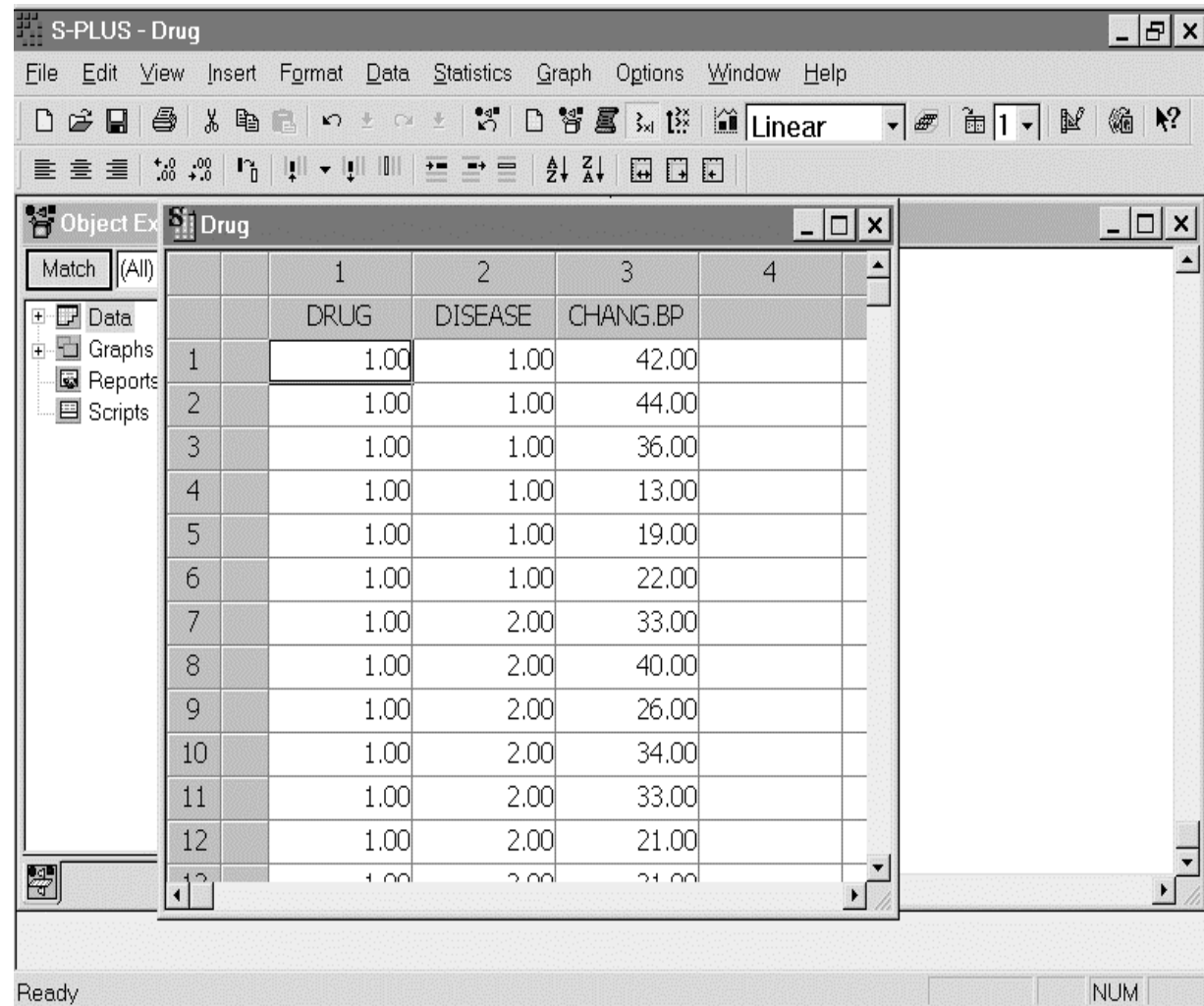
## Step 2



## Step 3



Result



S-PLUS - Drug

File Edit View Insert Format Data Statistics Graph Options Window Help

Linear 1

Object Ex Drug

Match (All)

Data  
Graphs  
Reports  
Scripts

	1	2	3	4
	DRUG	DISEASE	CHANG.BP	
1	1.00	1.00	42.00	
2	1.00	1.00	44.00	
3	1.00	1.00	36.00	
4	1.00	1.00	13.00	
5	1.00	1.00	19.00	
6	1.00	1.00	22.00	
7	1.00	2.00	33.00	
8	1.00	2.00	40.00	
9	1.00	2.00	26.00	
10	1.00	2.00	34.00	
11	1.00	2.00	33.00	
12	1.00	2.00	21.00	

Ready NUM

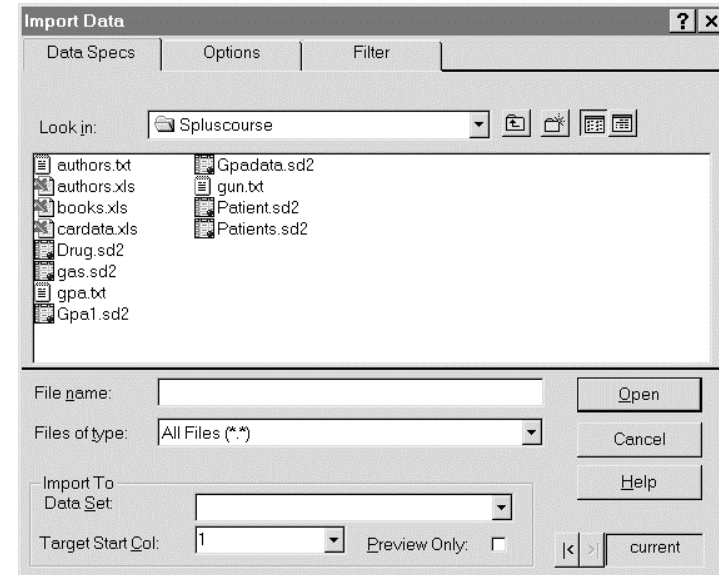
# Data Specs Page:

**File Name:** Specify the file name

**Files of type:** Select the form of the file you want to import.

**Data Frame:** Specify the data frame to contain the imported data.

**Target Start Col.:** Specify the starting column for the imported data. The start column can be any column in an existing data frame or a new data frame.



**Preview Only:** If you choose Preview Only the data is read entirely as a character columns into a data frame called PREVIEW. This allows you to view all of the data regardless of the data type. 92

# Options Page:

**Name Row:** If the file you are importing contains names for the columns of data, S-PLUS can use these names as column names.

**Start Column:** Specify the location in the file to begin reading from.

**End Column:** Specify the location of the last column in the file to read.

**Start Row:** Specify which row in the data file to begin reading from.

**End Row:** Specify the location of the last row to read.

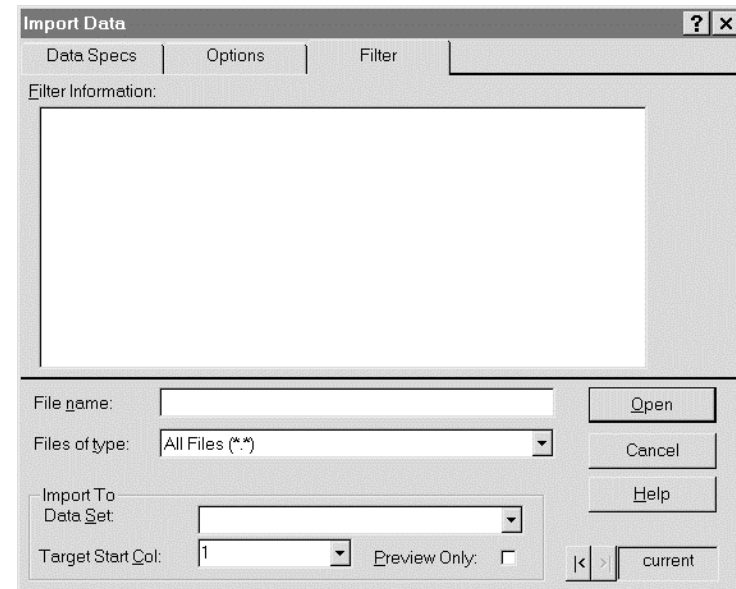
The screenshot shows the 'Import Data' dialog box with the 'Options' tab selected. The 'Data Specs' tab is also visible. The 'Options' tab contains the following fields and controls:

- Import Options (ASCII):**
  - Name Row: [ ]
  - Name Col: [ ]
  - Start Col: [1]
  - End Col: [END]
  - Start Row: [1]
  - End Row: [END]
  - Import Text as Factors: [Auto]
  - Import Value Labels as Numbers: [ ]
  - Delimiters: [ , \t ]
  - Separate Delimiters: [ ]
  - Dates as Doubles: [ ]
- ASCII Column Formats (Optional):**
  - Column Names: [ ]
  - Format String: [ ]
  - E.g., if Column Names named a character column and a numeric column, and the input file had an extra column to be skipped, the Format String might resemble this: "%s %\* %f"
- File name:** [ ]
- Files of type:** [All Files (\*.\*)]
- Import To Data Set:** [ ]
- Target Start Col:** [1]
- Preview Only:** [ ]
- Buttons:** Open, Cancel, Help, <, >, current

**Delimiter:** Specify all characters used to separate elements in an ASCII file. Commas, spaces and tabs are the default delimiters. Carriage returns and line feed are not allowed because they terminate each row.

# Filter Page:

It allows us to subset the data we import. By specifying a query. Also by using the filter page you can gain additional functionality as taking a random sample from the data.



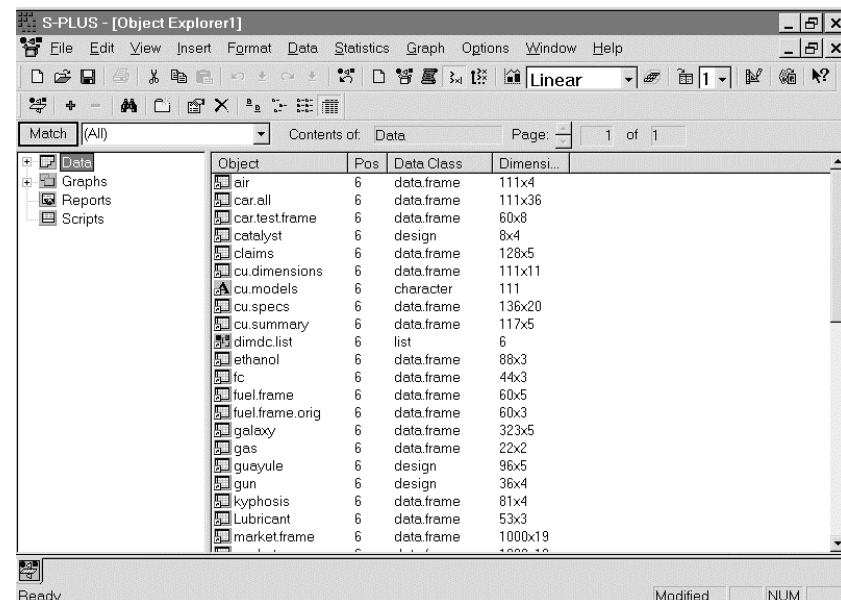
## Note:

The filter page for the file import dialogue is ignored if the File type is set to

- ASCII
- Formatted ASCII
- S-Plus Transport

# Viewing Data with the Data Window

- List the data frames in your S-Plus session
  - Click on the + connected to data frame, in the left pane of the Object Browser
  - Double-click on the data you want to view (say **air** data frame)



S-PLUS - [air]

File Edit View Insert Format Data Statistics Graph Options Window Help

Linear 1

	1	2	3	4	5	6	7
	ozone	radiation	temperature	wind			
1	3.45	190.00	67.00	7.40			
2	3.30	118.00	72.00	8.00			
3	2.29	149.00	74.00	12.60			
4	2.62	313.00	62.00	11.50			
5	2.84	299.00	65.00	8.60			
6	2.67	99.00	59.00	13.80			
7	2.00	19.00	61.00	20.10			
8	2.52	256.00	69.00	9.70			
9	2.22	290.00	66.00	9.20			
10	2.41	274.00	68.00	10.90			
11	2.62	65.00	58.00	13.20			
12	2.41	334.00	64.00	11.50			
13	3.24	307.00	66.00	12.00			
14	1.82	78.00	57.00	18.40			
15	3.11	322.00	68.00	11.50			

Ready NUM

# Editing Data

- Edit column names
- Creating new variables
- Select rows
- Select columns
- Select sub-sections of the data frame.

# Editing Column Names

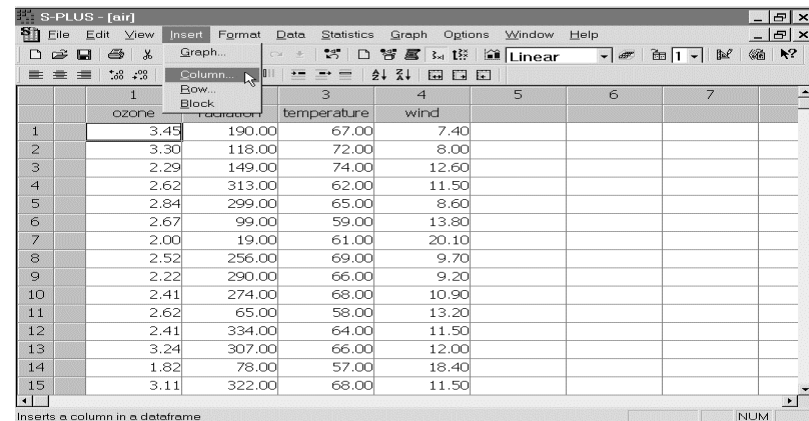
- To change column names
  - double Click on the column name to highlight it
  - type in the new name
- Notes:
  - Column Names can contain letters, numbers and periods
  - S-Plus is case sensitive so a column name “Age” is different than “age”
  - They must start with a letter
  - S-Plus function names can not be used as column names
  - A column name can have up to 75 characters
  - A column name can’t use spaces or underscores

# Creating New Variables

- Open the data frame you need to insert a new column to.
- From the “**Main Menu**”
  - Click Insert
  - Click Column
  - The Insert Column dialog will appear
    - Enter a name for the new column
    - Go to Fill Expression text field and type in the formula or value you want the column to have
    - Go to the Target text field to specify the target data frame
    - In the Column Type specify the data type for the inserted column
    - Click Ok

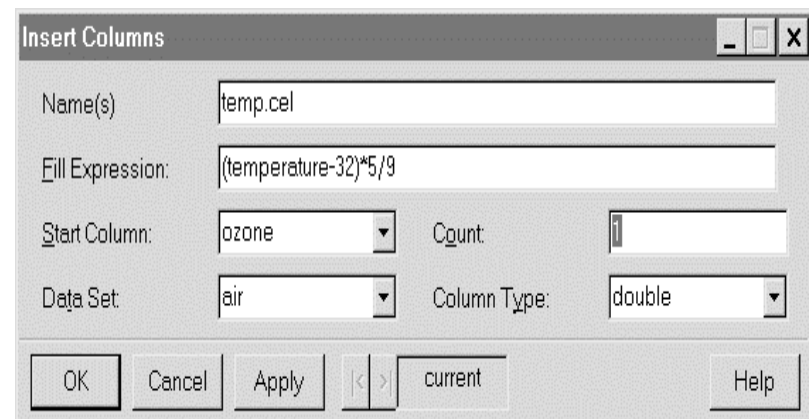
# Example: Creating New Variables

- For the data frame **air**, Create the variable “**temp.cel**” where
  - temp.cel: is the temperature in degrees Celsius.
    - $(\text{temperature} - 32) * 5/9$



The screenshot shows the S-PLUS [air] window. The menu bar includes File, Edit, View, Insert, Format, Data, Statistics, Graph, Options, Window, and Help. The 'Insert' menu is open, showing options like Graph..., Column..., Row..., and Block. The data table below has columns labeled 1, 2, 3, 4, 5, 6, and 7. The first three columns contain data for 'ozone', 'temperature', and 'wind' respectively. The data is as follows:

	1	2	3	4	5	6	7
1	3.45	190.00	67.00	7.40			
2	3.30	118.00	72.00	8.00			
3	2.29	149.00	74.00	12.60			
4	2.62	313.00	62.00	11.50			
5	2.84	299.00	65.00	8.60			
6	2.67	99.00	59.00	13.80			
7	2.00	19.00	61.00	20.10			
8	2.52	256.00	69.00	9.70			
9	2.22	290.00	66.00	9.20			
10	2.41	274.00	68.00	10.90			
11	2.62	65.00	58.00	13.20			
12	2.41	334.00	64.00	11.50			
13	3.24	307.00	66.00	12.00			
14	1.82	78.00	57.00	18.40			
15	3.11	322.00	68.00	11.50			



The 'Insert Columns' dialog box is shown with the following fields and values:

- Name(s): temp.cel
- Fill Expression: (temperature-32)\*5/9
- Start Column: ozone
- Count: 1
- Data Set: air
- Column Type: double

Buttons at the bottom include OK, Cancel, Apply, a navigation button with left and right arrows, a 'current' button, and a Help button.

Result

S-PLUS - [air]

File Edit View Insert Format Data Statistics Graph Options Window Help

Linear 1

	1	2	3	4	5	6	7
	ozone	radiation	temperature	wind	temp.cel		
1	3.45	190.00	67.00	7.40	19.44		
2	3.30	118.00	72.00	8.00	22.22		
3	2.29	149.00	74.00	12.60	23.33		
4	2.62	313.00	62.00	11.50	16.67		
5	2.84	299.00	65.00	8.60	18.33		
6	2.67	99.00	59.00	13.80	15.00		
7	2.00	19.00	61.00	20.10	16.11		
8	2.52	256.00	69.00	9.70	20.56		
9	2.22	290.00	66.00	9.20	18.89		
10	2.41	274.00	68.00	10.90	20.00		
11	2.62	65.00	58.00	13.20	14.44		
12	2.41	334.00	64.00	11.50	17.78		
13	3.24	307.00	66.00	12.00	18.89		
14	1.82	78.00	57.00	18.40	13.89		
15	3.11	322.00	68.00	11.50	20.00		

Ready NUM

# Selecting Rows

- **To select rows that are contiguous next to each other)**
  - Click and hold the left mouse button on the number of the first row that you are choosing
  - Drag the mouse across the row number you are selecting. Everything that you selected should be highlighted.

OR

- Click on the first row number for the row you want to select.
  - Shift-click on the last row number for the rows you want to select. Everything that you selected should be highlighted.
- **To select discontinuous rows**
  - Click on the number for the first row you want to select.
  - CTRL-click the remaining row numbers you are selecting.

# Selecting Columns

- **To select columns that are contiguous next to each other)**
  - Click and hold the left mouse button on the first column name (or number) that you are choosing
  - Drag the mouse across the column name (or number) you are selecting. Everything that you selected should be highlighted.

OR

- Click on the first column name (or number) of the column you want to select.
  - Shift-click on the last column name (or number) of the columns you want to select. Everything that you selected should be highlighted.
- **To select discontinuous rows**
  - Click on the first column name (or number) you want to select.
  - CTRL-click the remaining columns (numbers) you are selecting.

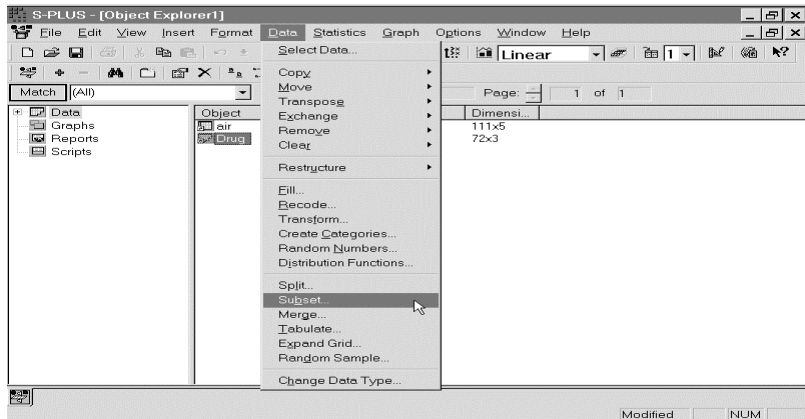
# Selecting Subsets (rows & columns) of Data

- To do so you need to use combinations of the following techniques
  - Click-hold-drag technique
  - Shift-click technique
  - CTRL-click technique

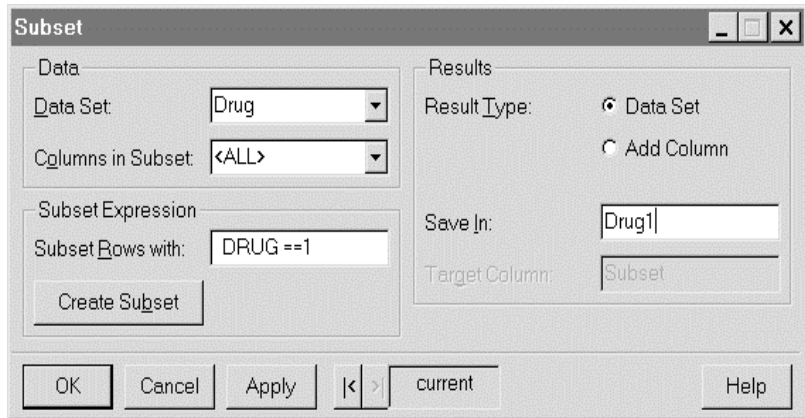
# Removing Observations

- Create a data set from the Drug data set that contains only the data when  $DRUG = 1$ ?
- To do so
  - select **Data** from the Main Menu
  - select **subset**
  - in the subset dialog
    - in Subset rows with fill **DRUG= = 1**
    - in **Save As** fill **Drug1**

Step 1



Step 2



Result

The screenshot shows the S-PLUS [Drug1] window displaying a data table with 20 rows and 7 columns. The columns are labeled 1 through 7, with the first three containing data: DRUG, DISEASE, and CHANG.BP. The data is as follows:

	1	2	3	4	5	6	7
	DRUG	DISEASE	CHANG.BP				
6	1.00	1.00	22.00				
7	1.00	2.00	33.00				
8	1.00	2.00	40.00				
9	1.00	2.00	26.00				
10	1.00	2.00	34.00				
11	1.00	2.00	33.00				
12	1.00	2.00	21.00				
13	1.00	3.00	31.00				
14	1.00	3.00	-3.00				
15	1.00	3.00	19.00				
16	1.00	3.00	25.00				
17	1.00	3.00	25.00				
18	1.00	3.00	24.00				
19							
20							

# Merging Data

- Different Column Names and Matched Rows
- Different Column Names and Not-Matched Rows

## Different Column Names and Matched Rows

- Consider the two data sets authors.txt and books.xls
  - Import the two data sets into S-plus
  - Click **Data**
  - Click **Merge**
  - Type **authors** in the **Data Frame 1** text field, and type **books** in the **Data Frame 2** text field
  - Click the button for **Specified Cols.** In the **Row Matching, Match by** section

- Type or click/CTRL-click the names of the columns from the **Data Frame 1** data frame that correspond to the columns you are making equivalent to in the **Data Frame 2** data frame.
  - For authors, choose **FirstName** and **LastName**
  - For books, choose **AuthorFirstName** and **AuthorLastName**
- In the Save As text field use new.books as the name of the name of the new data frame
- Click OK.
- Only the matched values found in specified columns of **authors** and **books** will be merged in the final data frame.

## Step 1

	1	2	3	4	5
	FirstName	LastName	Age	Income	Home
1	Lorne	Green	82.00	200000.00	California
2	Loren	Jaye	40.00	40000.00	Washington
3	Robin	Green	45.00	25000.00	Washington
4	Robin	Howe	2.00	0.00	Alberta
5	Billy	Jaye	40.00	27500.00	Washington
6					
7					
8					

	1	2	3	4
	AuthorFirstName	AuthorLastName	Book	
1	Lorne	Green	Bonanza	
2	Loren	Jaye	Midwifery	
3	Loren	Jaye	Gardening	
4	Loren	Jaye	Perennials	
5	Robin	Green	Who_dun_it	
6	Rich	Calaway	Splus	
7				
8				

## Step 2

**Merge Two Data Frames**

Data Frames  
 Data Frame 1: authors  
 Data Frame 2: books

Row Matching  
 Match by:  
☐ All Common Cols  
☐ Row Names  
☒ Specified Cols

Columns in D.F. 1: FirstName, LastName  
 Columns in D.F. 2: AuthorFirstName, AuthorLastName

Include Non-Matched Rows in:  
☐ Data Frame 1  
☐ Data Frame 2

Suffix for Non-Matching Common Cols:  
 Data Frame 1: .1  
 Data Frame 2: .2

Results  
 Save As: new.books  
☐ Show in Data Window

OK Cancel Apply < > current Help

## Result

	1	2	3	4	5	6
	FirstName	LastName	Age	Income	Home	Book
1	Loren	Jaye	40.00	40000.00	Washington	Midwifery
2	Loren	Jaye	40.00	40000.00	Washington	Gardening
3	Loren	Jaye	40.00	40000.00	Washington	Perennials
4	Lorne	Green	82.00	1200000.00	California	Bonanza
5	Robin	Green	45.00	25000.00	Washington	Who_dun_it
6						
7						
8						
9						

## Different Column Names and Non-Matched Rows

- Consider the two data sets authors.txt and books.xls. Here we want to merge all of the data in both data frames
  - Click **Data**
  - Click **Merge**
  - Type **authors** in the **Data Frame 1** text field, and type **books** in the **Data Frame 2** text field
  - Click the button for **Specified Cols.** In the **Row Matching, Match by** section

- Type or click/CTRL-click the names of the columns from the **Data Frame 1** data frame that correspond to the columns you are making equivalent to in the **Data Frame 2** data frame.
  - For authors, choose **FirstName** and **LastName**
  - For books, choose **AuthorFirstName** and **AuthorLastName**
- Click on the **Data Frame 1** and **Data Frame 2** boxes in the Include Non-Matched Rows in section of the dialog.
- In the Save As text field use **all.books** as the name of the name of the new data frame
- Click OK.
- S-Plus merges all columns, placing **NA** where no information is available.

## Step 1

	1	2	3	4	5
	FirstName	LastName	Age	Income	Home
1	Lorne	Green	82.00	200000.00	California
2	Loren	Jaye	40.00	40000.00	Washington
3	Robin	Green	45.00	25000.00	Washington
4	Robin	Howe	2.00	0.00	Alberta
5	Billy	Jaye	40.00	27500.00	Washington
6					
7					
8					

	1	2	3	4
	AuthorFirstName	AuthorLastName	Book	
1	Lorne	Green	Bonanza	
2	Loren	Jaye	Midwifery	
3	Loren	Jaye	Gardening	
4	Loren	Jaye	Perennials	
5	Robin	Green	Who_dun_it	
6	Rich	Calaway	Splus	
7				
8				

## Step 2

**Merge Two Data Frames**

Data Frames  
 Data Frame 1: authors  
 Data Frame 2: books

Row Matching  
 Match by:  
☐ All Common Cols  
☐ Row Names  
☒ Specified Cols

Columns in D.F. 1: FirstName, LastName  
 Columns in D.F. 2: AuthorFirstName, AuthorLastName

Include Non-Matched Rows in:  
☒ Data Frame 1  
☒ Data Frame 2

Suffix for Non-Matching Common Cols:  
 Data Frame 1: .1  
 Data Frame 2: .2

Results:  
 Save As: all.books  
☐ Show in Data Window

OK Cancel Apply |< >| current Help

## Result

	1	2	3	4	5	6
	AuthorFirstName	AuthorLastName	Book	Age	Income	Home
1	Billy	Jaye	NA	40.00	27500.00	Washington
2	Loren	Jaye	Midwifery	40.00	40000.00	Washington
3	Loren	Jaye	Gardening	40.00	40000.00	Washington
4	Loren	Jaye	Perennials	40.00	40000.00	Washington
5	Lorne	Green	Bonanza	82.00	200000.00	California
6	Rich	Calaway	Splus	NA	NA	NA
7	Robin	Green	Who_dun_it	45.00	25000.00	Washington
8	Robin	Howe	NA	2.00	0.00	Alberta
9						
10						
11						
12						

# Graphics In S-PLUS

# **Objectives:**

- Anatomy of the Graph Sheet
- Creating Graphs
  - Univariate Plots
  - 2D Plots
  - 3D Plots
- Modifying graphs
- Combining multiple plots on one graph or on one page
- Saving Graphs
- Plotting Subsets of Rows
- Using Text as Symbols
- Brush and Spin

# **Anatomy of Graph Sheet**

- ***Graph Sheet***: the actual page the plot is drawn on.
- ***Graph Area***: the area outside the plot axis but within the Graph Sheet.
- ***Plot Area***: the area of the plot including the axis.
- ***Plot***: the actual points that contains the plot.

# **Creating Graphs**

- Open the Data Window.
- Select the data in the data window/object browser.
- Open the appropriate plot Palette.
- Select a plot.

# Univariate Plots

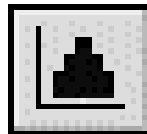
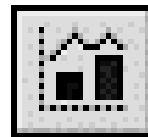
They are used to assess the distribution of the data like

- Histograms
- Boxplots
- Quantile plots

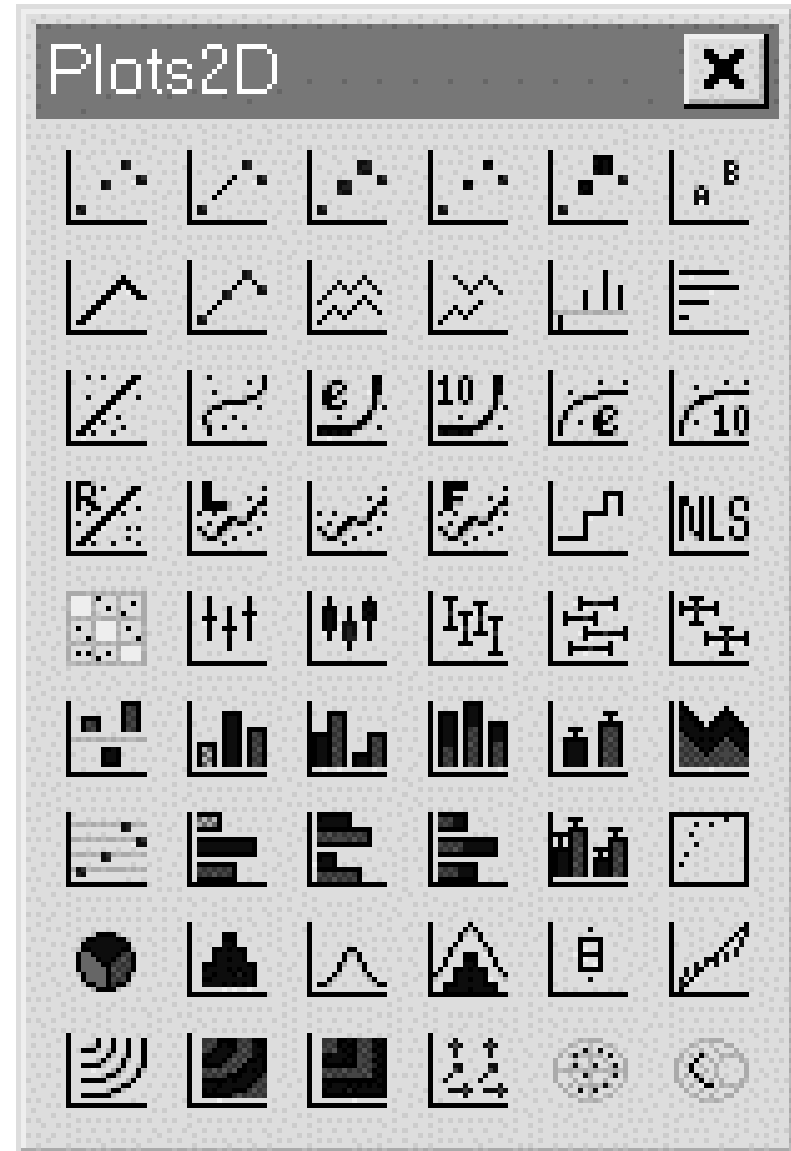
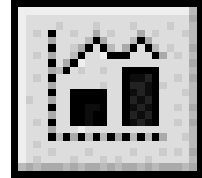
# Histograms:

- They are used to assess the distribution of data and visualize potential extreme values of the data.
- They are created by dividing the range of the data into a number of cells or bins, with each of the bins containing a vertical bar whose height reflects the number of observations in that bin.

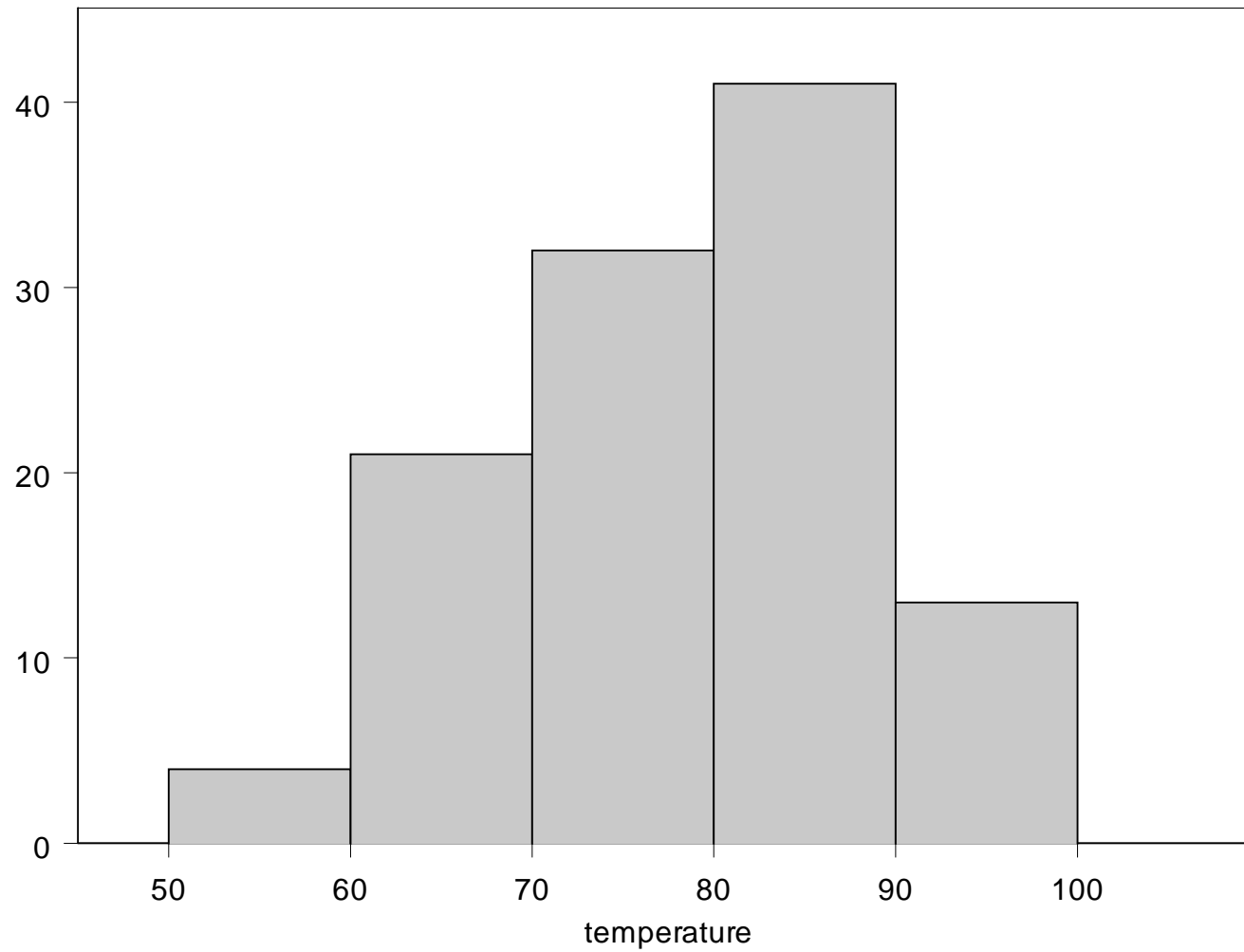
- From the **air** data frame in S-Plus.
  - Create a histogram for the temperature variable.
  - Modify Plot Attributes
  - Add Title to the Histogram
- **Creating Histogram**
  - Open the Data Window by double clicking on air in the object browser page.
  - Select the data in the Data window by clicking on the “temperature” column.
  - Open the “2D plot palette”.
  - Select “Histogram”.



# \* The 2D Platte



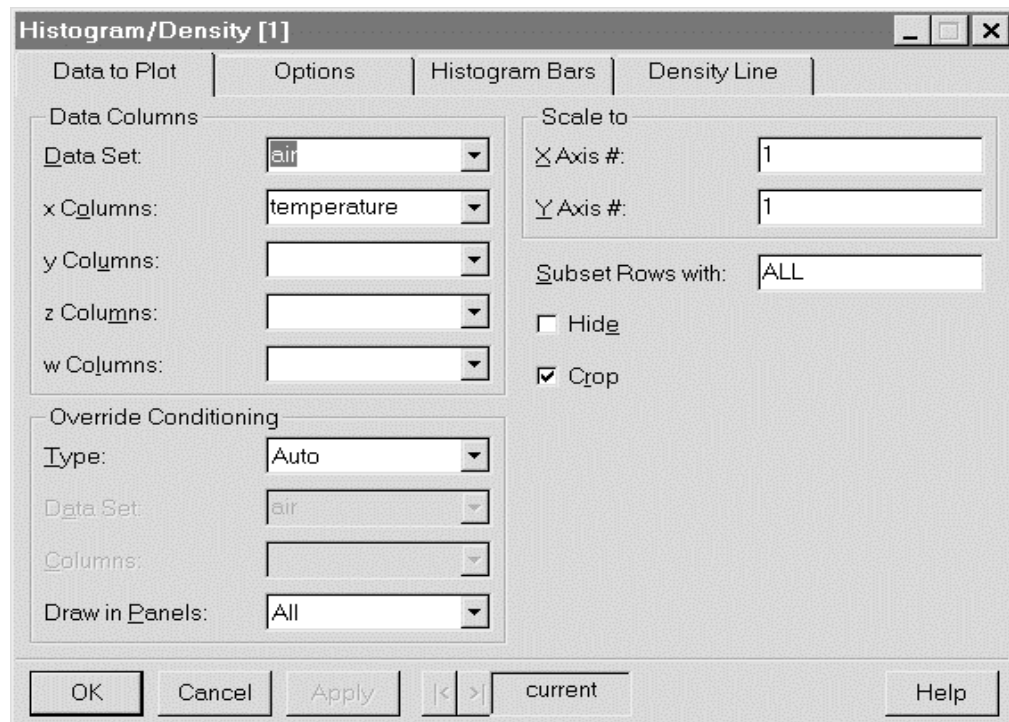
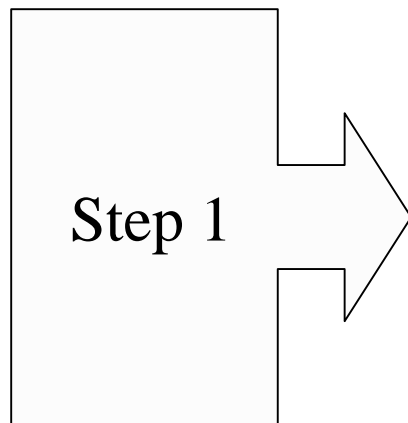
The result is



# Modifying Plot Attributes

- Open the Histogram/Density dialog
  - By Double Click on the histogram
- Change the plot attributes
  - Click the **Histogram Bars** tab
  - Change the **Fill Color: Lt. Magenta**
  - Change the **Fill Pattern: \*\*\*\*\***
  - Change **Pattern Color: Black**
  - Change **Border/Color: Black**
  - Change **Border/Weight: 2**

- Click the check box for **Draw Bars**.
- Click the **Options** tab
- In the Bars section change the number of Bars to 12.
- Click Ok



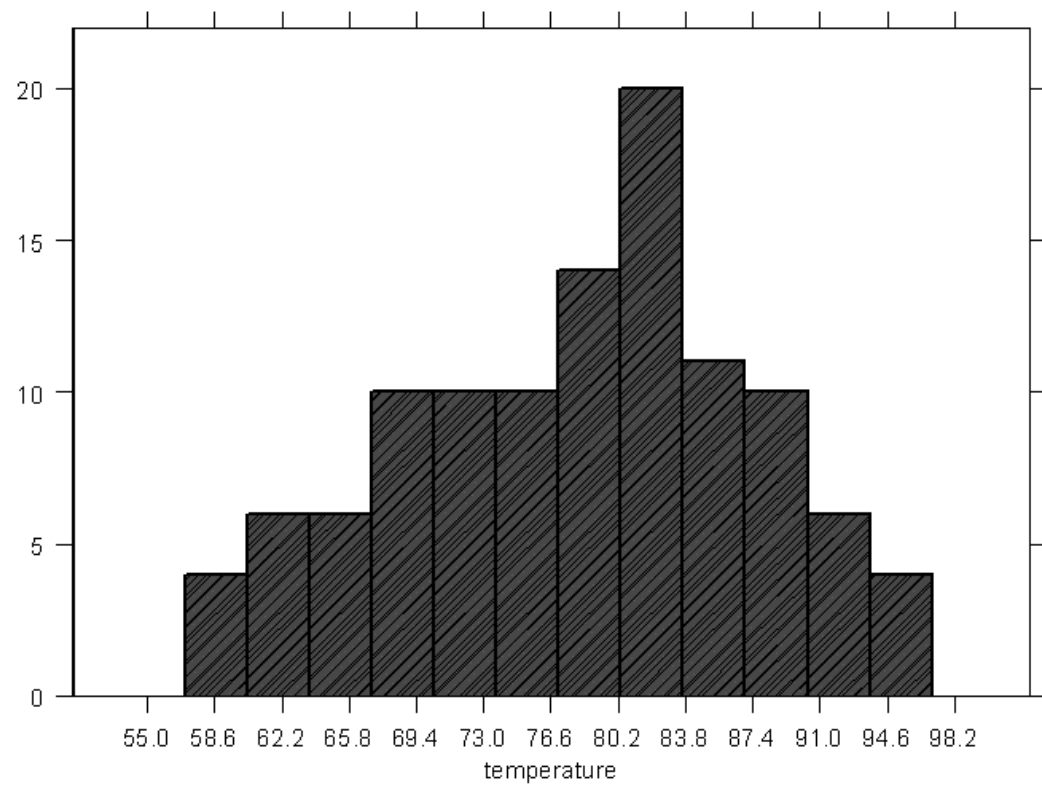
Step 2

The dialog box is titled "Histogram/Density [1]". It has four tabs: "Data to Plot", "Options", "Histogram Bars", and "Density Line". The "Histogram Bars" tab is selected. Under the "Fill" section, "Fill Color" is set to "Lt Magenta", "Fill Pattern" is set to a diagonal line pattern, and "Pattern Color" is set to "Black". Under the "Border" section, "Style" is set to a solid line, "Color" is set to "Black", and "Weight" is set to "2". There are two checked checkboxes: "Draw Bars" and "Draw Histograms". At the bottom, there are buttons for "OK", "Cancel", "Apply", navigation arrows, "current", and "Help".

Step 3

The dialog box is titled "Histogram/Density [1]". It has four tabs: "Data to Plot", "Options", "Histogram Bars", and "Density Line". The "Density Line" tab is selected. Under the "Histogram Specs" section, "Contin./Integer" is set to "Continuous" and "Output Type" is set to "Counts". Under the "Data Range" section, "Lower Bound" and "Upper Bound" are both set to "Auto". Under the "Bars" section, "Interval Width" is set to "Auto" and "Number of Bars" is set to "12". Under the "Density Options" section, "Window" is set to "Gaussian", "Num Output Pts" is set to "Auto", "Window Width" is set to "Auto", "From" is set to "Auto", "To" is set to "Auto", and "Cut" is set to "Auto". There is an unchecked checkbox for "Remove NA's". At the bottom, there are buttons for "OK", "Cancel", "Apply", navigation arrows, "current", and "Help".

Result



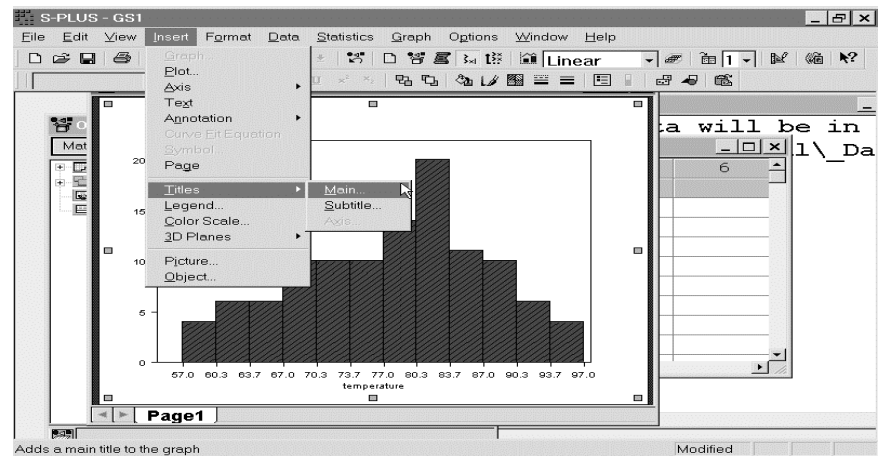
# Saving the Graph

- Make sure that the graph window is active.
- From the **File menu** do **Save As**
- Enter the name of the file you need to save the graph into. It should have the extension **sgr**.
- Hit **OK**. The graph is saved now.

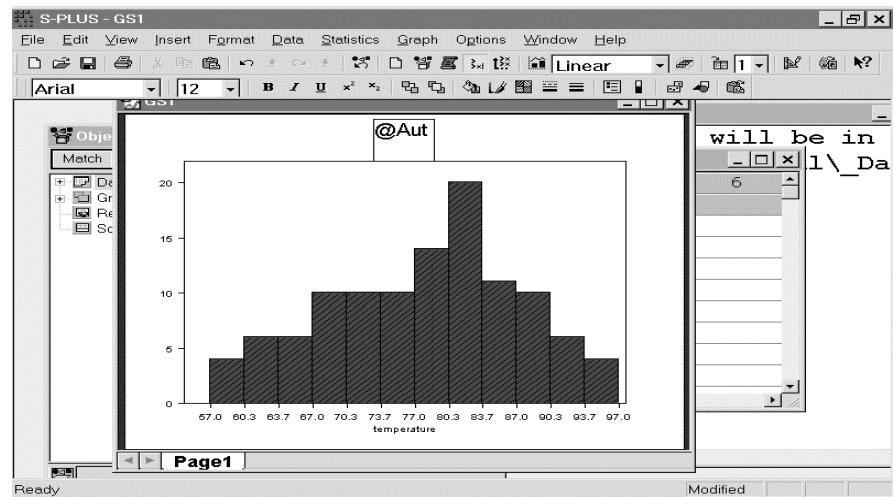
# Adding Titles

- You could add a main title or a subtitle
- From the **Main Menu**
  - Click Insert
  - Click Titles
  - Click Main or Subtitle
  - Type “Histogram Plot for Temperature”
  - Click outside the edit box to accept the title text

Step 1

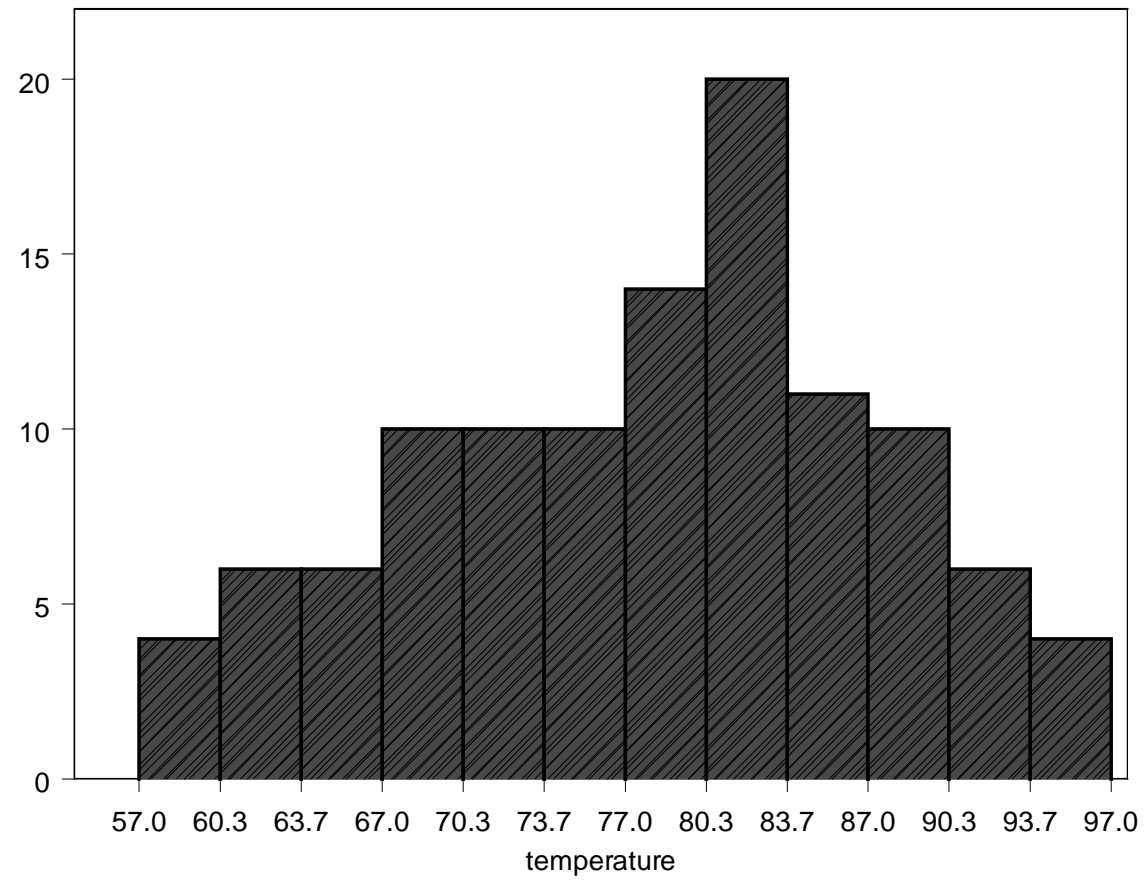


Step 2



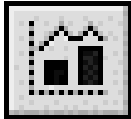
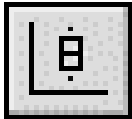
**Result**

Histogram for Temperature

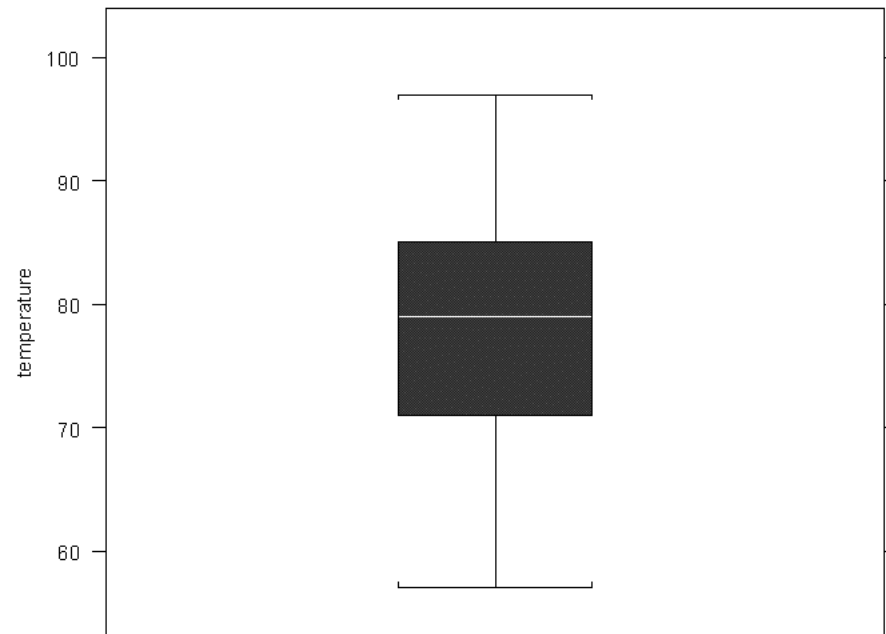


# Boxplots

- It is a plot that shows the five-number summary (minimum, 1st quartile, median, 3rd quartile, maximum)
  - The ends of the box are the 1st and 3rd quartile.
  - A line is drawn at the median.
  - Whiskers extend out to the minimum and the maximum value

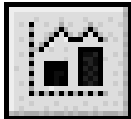
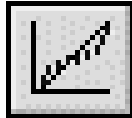
- From the **air** data frame in S-Plus.
  - Create a BoxPlot for the temperature variable.
- **Creating BoxPlot**
  - Open the Data Window by double clicking on air in the object browser page.
  - Select the data in the Data window by clicking on the “temperature” column.
  - Open the “2D plot palette”. 
  - Select “BoxPlot”. 

The result is

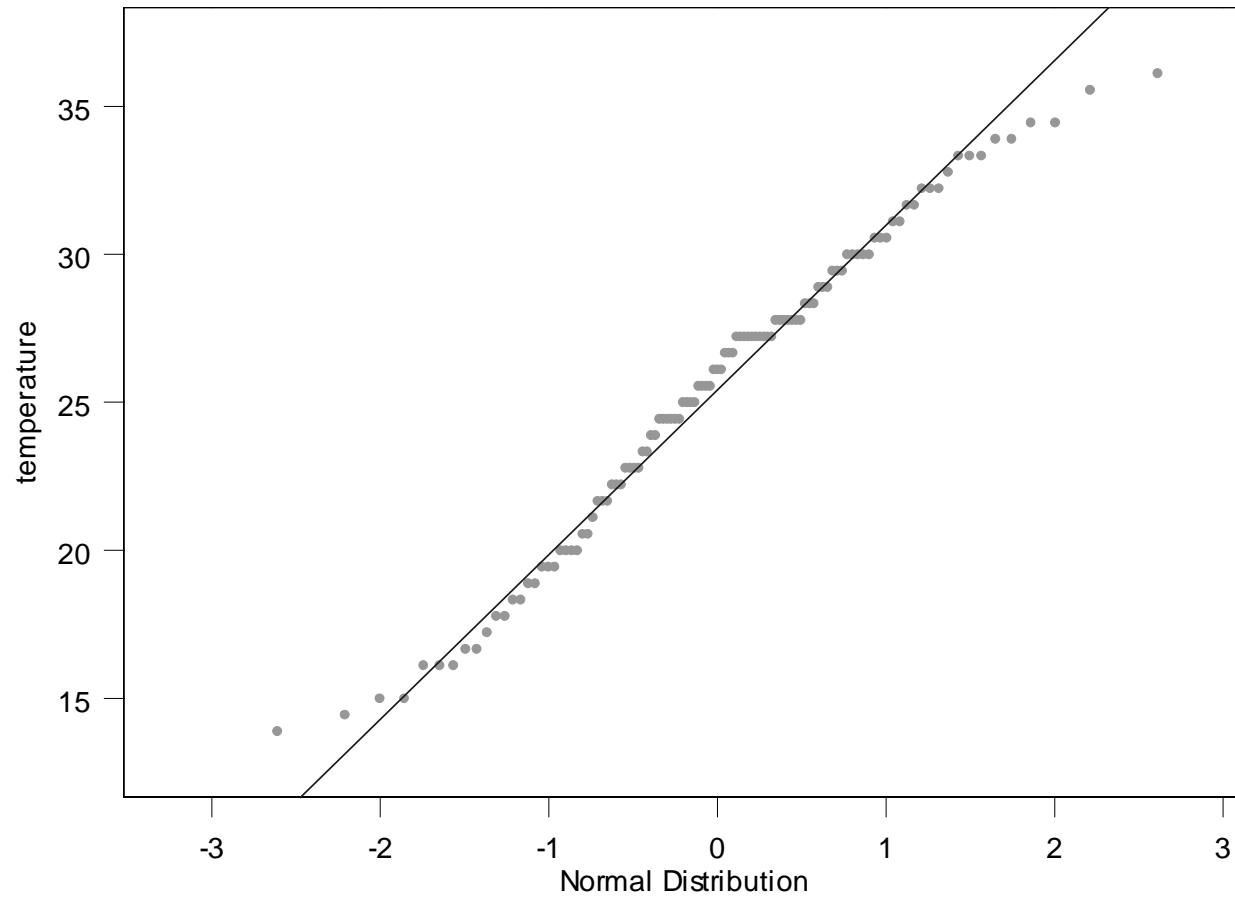


# Quantile Plots

- Those plots are useful for assessing data for normality or to other probability distribution, as well as to other data.

- From the **air** data frame in S-Plus.
  - Create a Quantile-Quantile Normal plot for the temperature variable.
- **Creating Quantile Plot**
  - Open the Data Window by double clicking on air in the object browser page.
  - Select the data in the Data window by clicking on the “temperature” column.
  - Open the “2D plot palette”. 
  - Select “QQ Normal with Line”. 

The result is

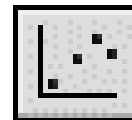
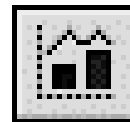


# 2D Scatter Plot

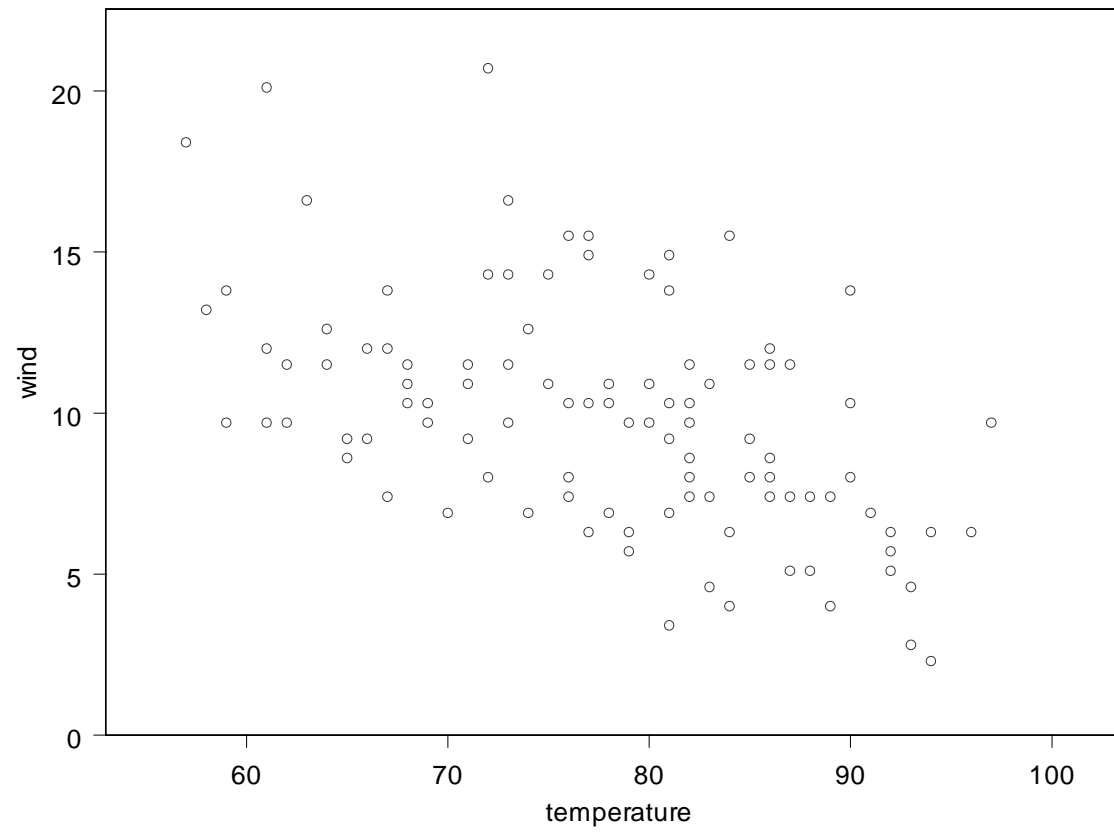
- From the **air** data frame in S-Plus.
  - Create a 2D scatter plot for the two variables temperature and wind (temperature on the x-axis).
- Then we are going to learn how to
  - Formatting Multiple Axis Titles Simultaneously
  - Formatting Axis Tick Ranges
  - Formatting Axis Tick Labels

- **Creating 2D Scatter Plot**

- Open the Data Window by double clicking on air in the object browser page.
- Select the data in the Data window.
  - By dragging across the column headers
  - Or, by CTRL-click on **temperature** and **wind**
  - Since we want temperature on the x-axis we select it first.
- Open the 2D plot palette.
- Select the 2D Scatter Plot



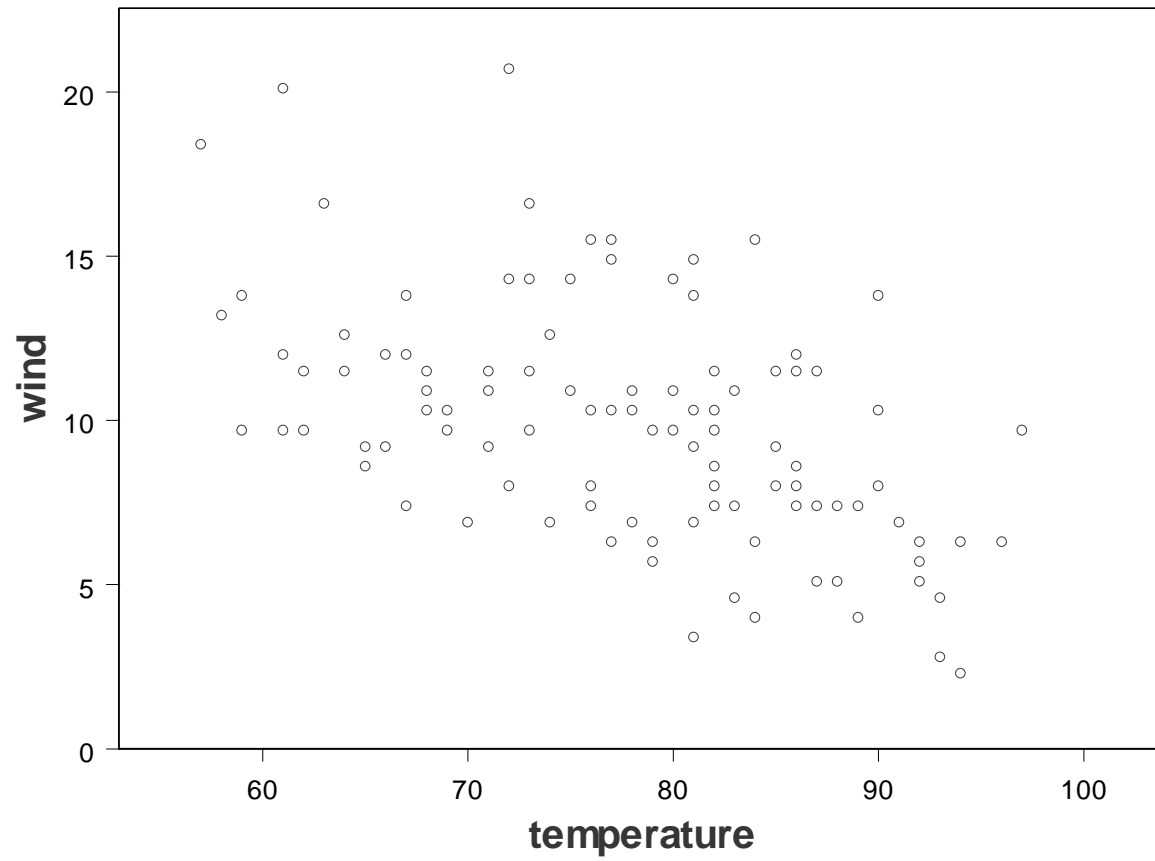
# The result:



## Formatting Multiple Axis Titles Simultaneously

- click on the X-axis Title
- CTRL-click on the Y-axis Title
- Using the “**Font Size**” Field on the tool bar, specify 24 for the font size
- Click the “**BOLD**” toolbar button
- Using the Line Color toolbar button, change the title colors to “**Lt. Red**”

Result



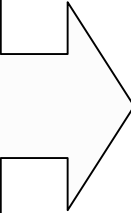
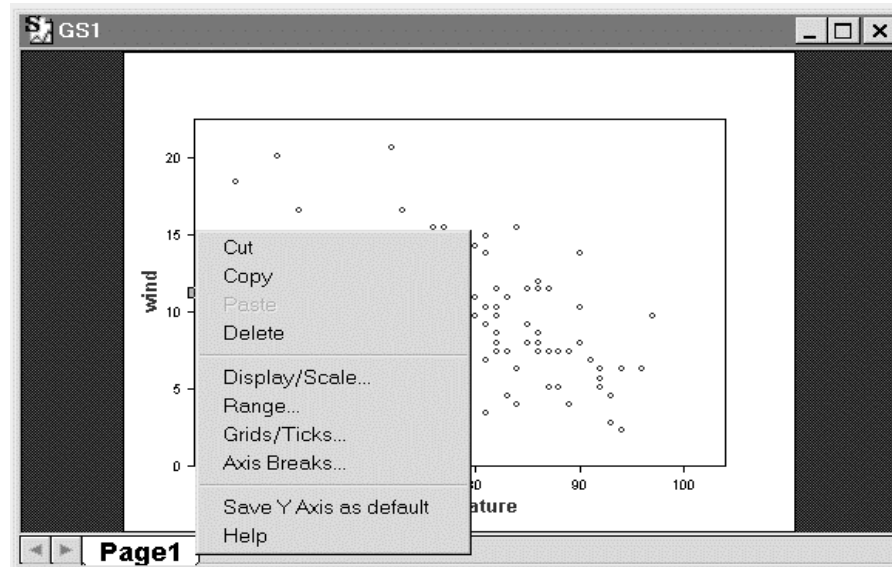
## **Formatting Axis Tick Ranges**

- S-Plus uses intelligent defaults to automatically calculate the placement of tick marks. However, you have a complete control over modifying them.

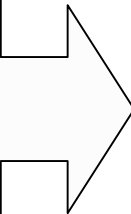
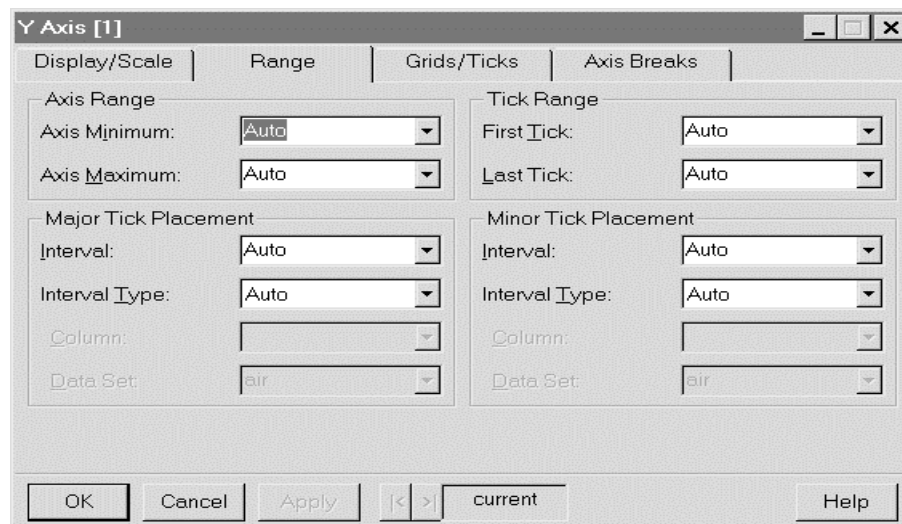
## **Example:**

- Right-click on the y axis and choose Range from the context menu
- Specify 0 for the Axis Minimum (here it is the case)
- Specify 30 for the Axis Maximum
- In the major tick Placement section, specify 3 for the interval and size for the Interval Type
- Click Ok

Step 1

Step 2

The screenshot shows the "Y Axis [1]" dialog box with the "Display/Scale" tab selected. The dialog box contains the following settings:

- Axis Range:**
  - Axis Minimum: Auto
  - Axis Maximum: Auto
- Major Tick Placement:**
  - Interval: Auto
  - Interval Type: Auto
  - Column: (empty)
  - Data Set: air
- Tick Range:**
  - First Tick: Auto
  - Last Tick: Auto
- Minor Tick Placement:**
  - Interval: Auto
  - Interval Type: Auto
  - Column: (empty)
  - Data Set: air

At the bottom of the dialog box, there are buttons for "OK", "Cancel", "Apply", and "Help". A "current" button is also present next to the "Apply" button.

Step 3

Y Axis [1]

Display/Scale | Range | Grids/Ticks | Axis Breaks

Axis Range

Axis Minimum: 0

Axis Maximum: 30

Major Tick Placement

Interval: 3

Interval Type: Size

Column:

Data Set: air

Tick Range

First Tick: Auto

Last Tick: Auto

Minor Tick Placement

Interval: Auto

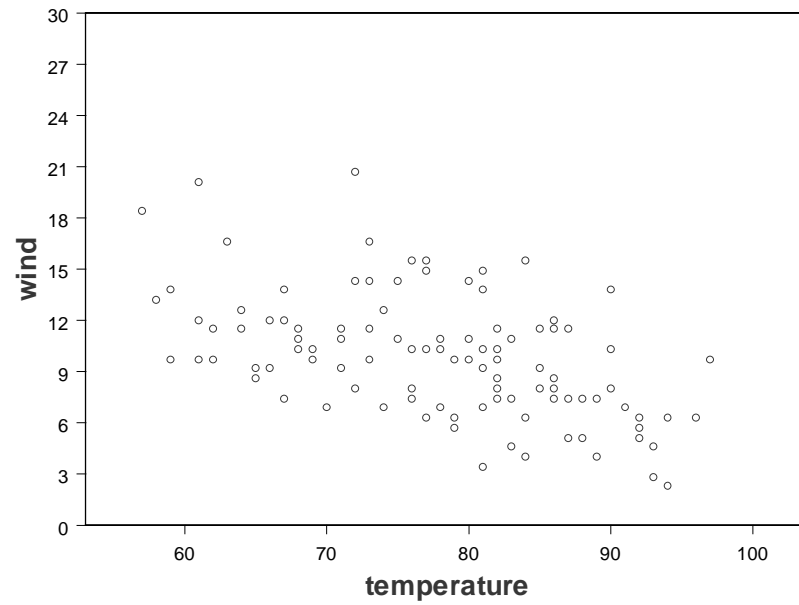
Interval Type: Auto

Column:

Data Set: air

OK Cancel Apply |< >| current Help

result



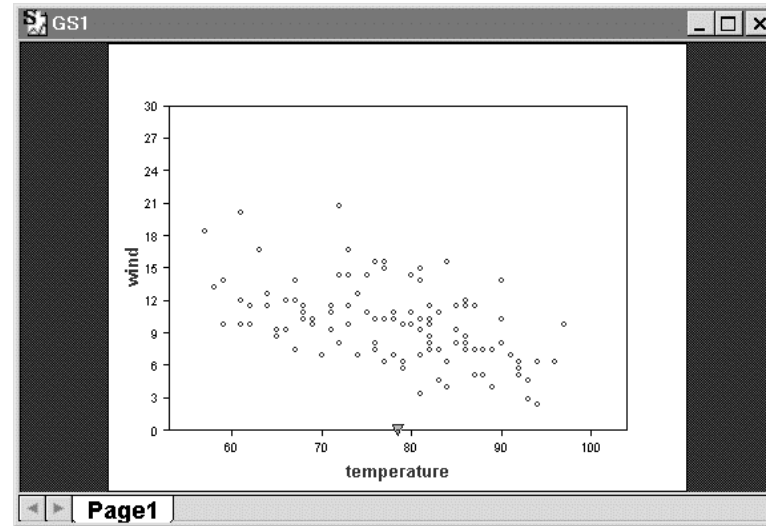
## **Formatting Axis Tick Labels**

- Axis tick labels are separate objects from the axes and can be selected and formatted independently.

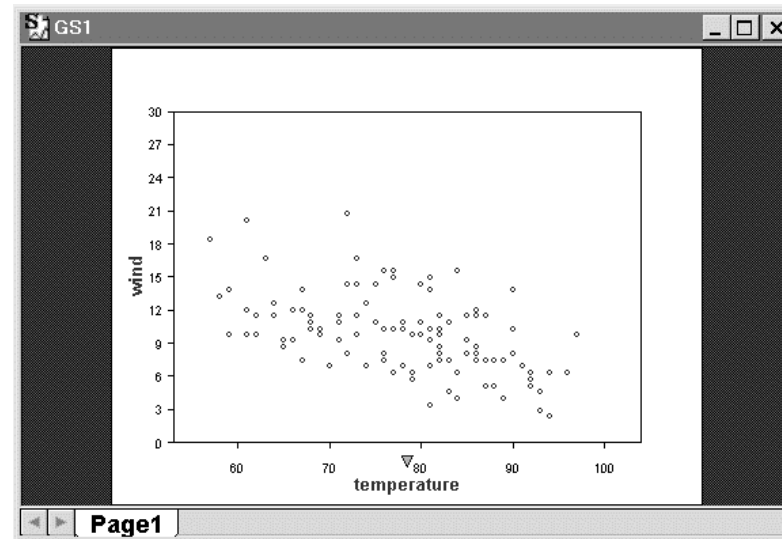
## **Example:**

- Click on the x axis label to select them.  
You will see a triangle selection knob.
- Drag the triangular selection knob to move the labels down a bit from the x axis.
- Right Click on the x axis labels and choose Font from the context menu
- Specify 20 in the Rotation field to rotate the label 20 degrees
- Click Ok.
- Click F9 to refresh the Screen.

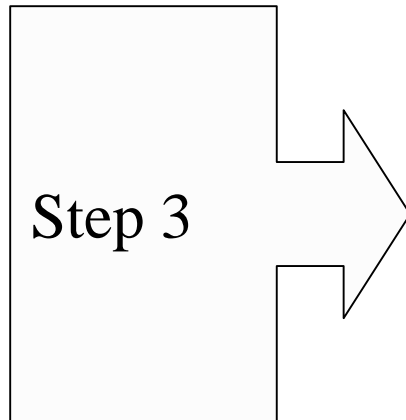
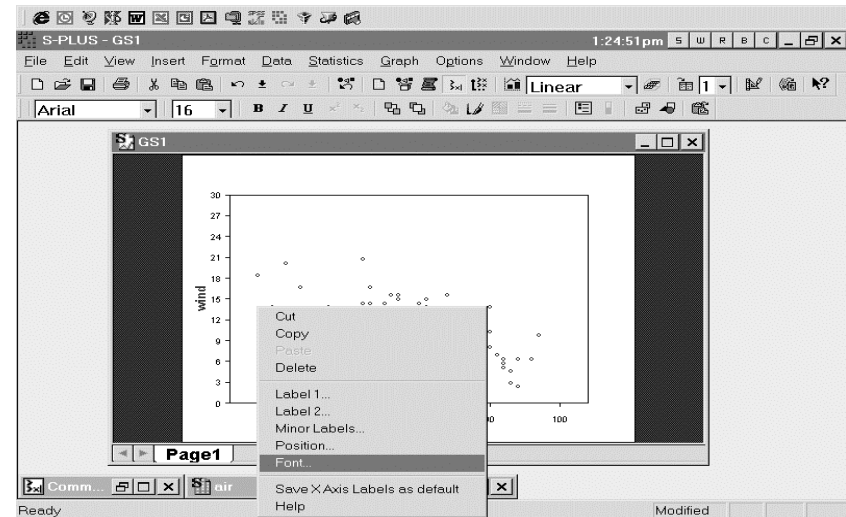
Step 1



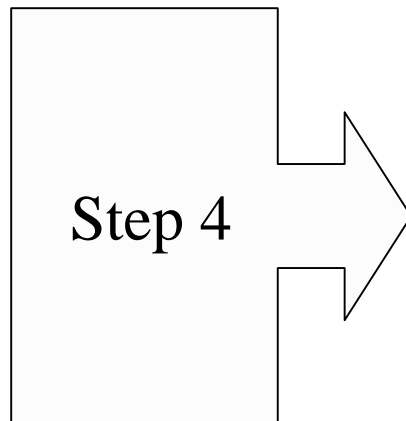
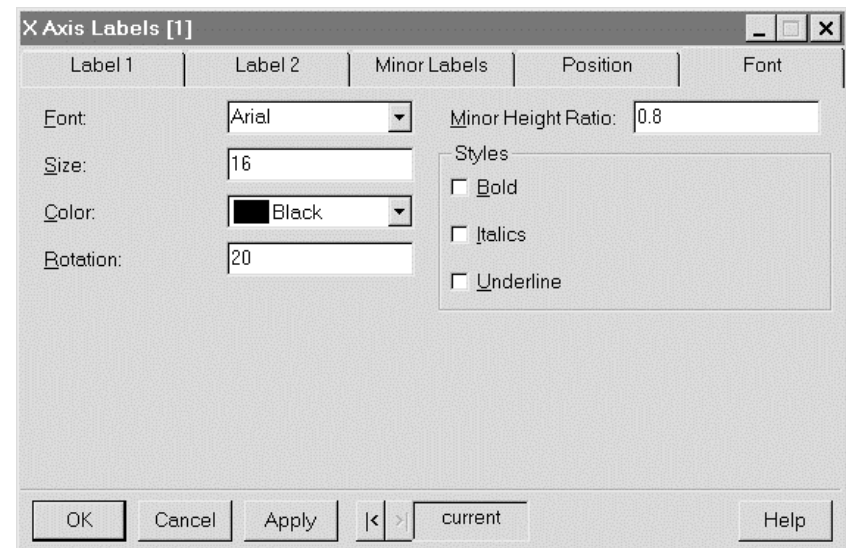
Step 2



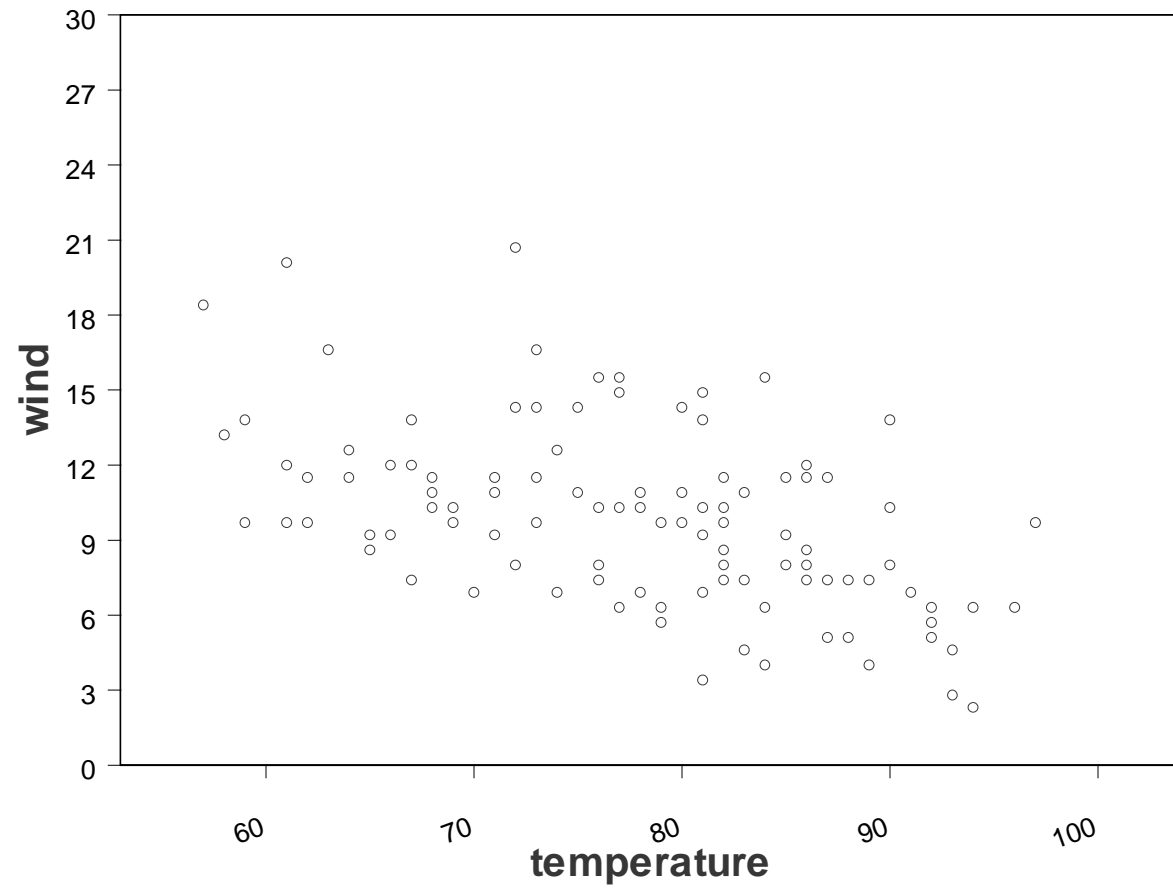
Step 3

Step 4

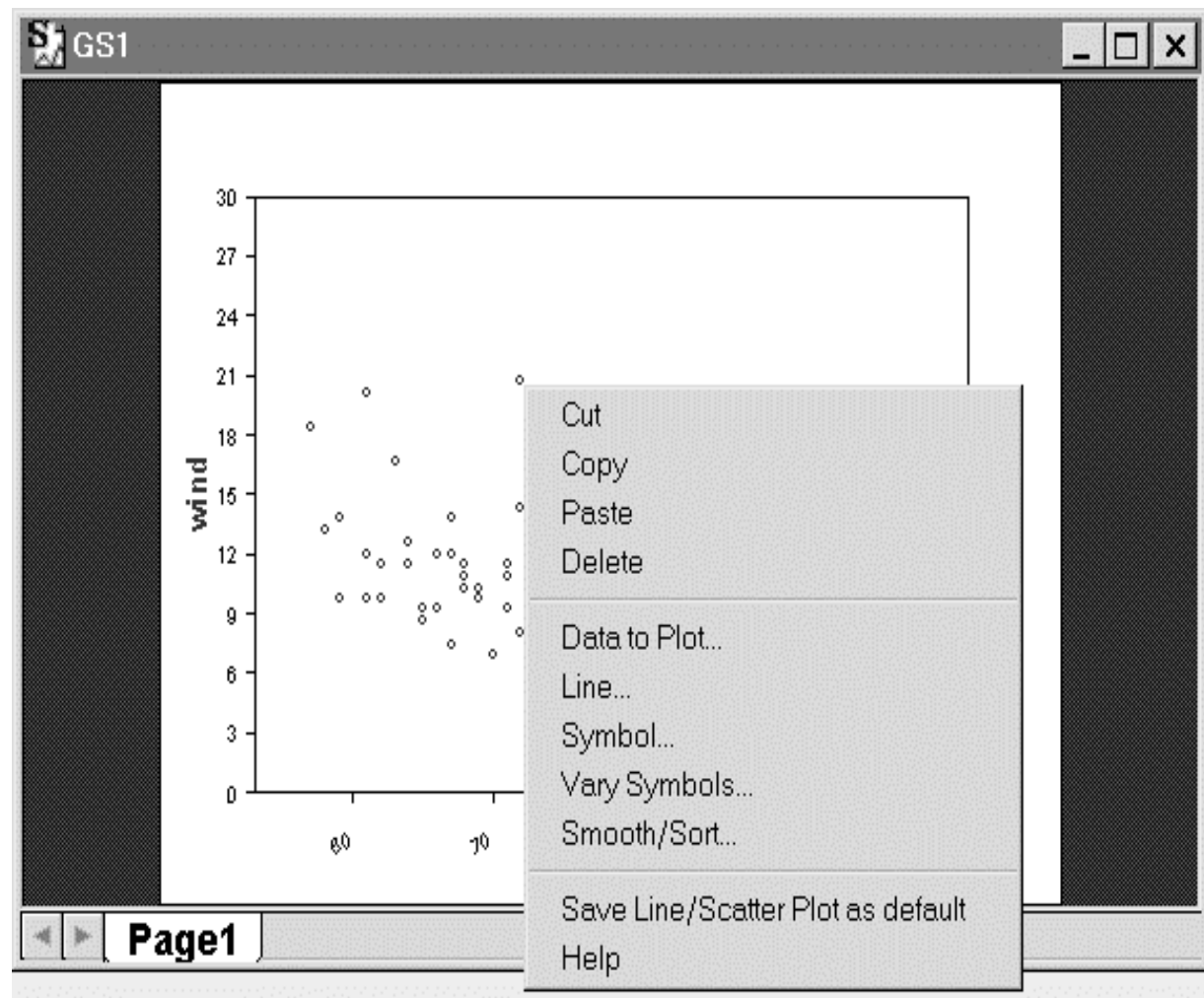
Result



# Changing Symbols and Colors

- **Example 1**
  - Open the Line/Scatter PLOT properties dialog by right-clicking on a symbol in the plot.
  - Click Symbol. The Line/Scatter Plot dialog opens.
  - Change the Symbol Style to **Circle, Solid**
  - Change the Symbol Color to **Blue**
  - Change the Symbol Height to **0.2**

Step 1



Step 2

Line/Scatter Plot [1]

Data to Plot | Line | **Symbol** | Vary Symbols | Smooth/Sort

Symbol

Style: ☐ Circle, Empty

Color:

Height:

Line Weight:

☐ Use Text As Symbol

Symbol Freq.:

Jitter Symbols:

Jitter Factor:

Text As Symbol

Text to Use:

Symbol Text:

Column:

Data Set:

Font:

☐ Bold

☐ Italics

☐ Underline

OK Cancel Apply |< >| current Help

Step 3

Line/Scatter Plot [1]

Data to Plot | Line | **Symbol** | Vary Symbols | Smooth/Sort

Symbol

Style: ☒ Circle, Solid

Color:

Height:

Line Weight:

☐ Use Text As Symbol

Symbol Freq.:

Jitter Symbols:

Jitter Factor:

Text As Symbol

Text to Use:

Symbol Text:

Column:

Data Set:

Font:

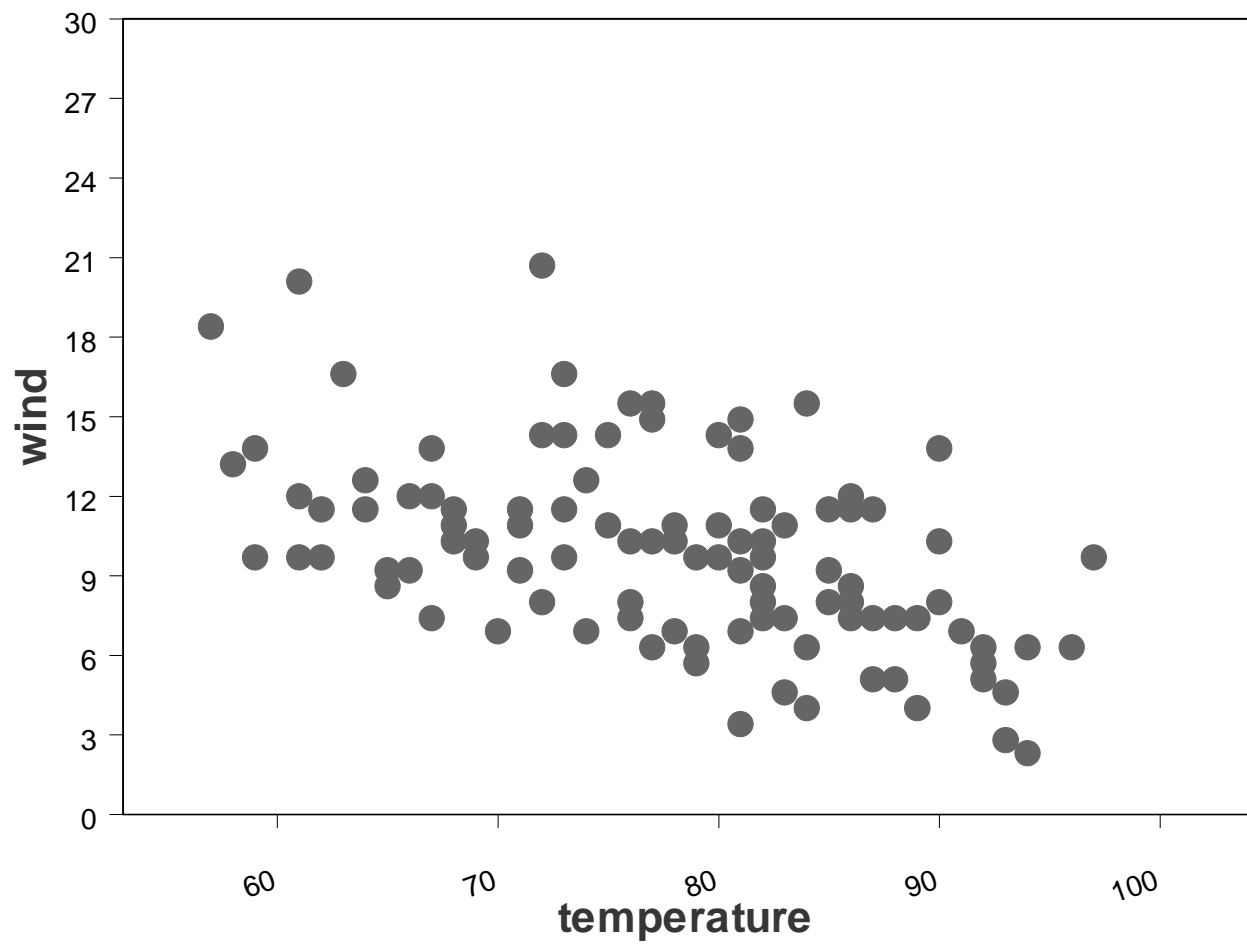
☐ Bold

☐ Italics

☐ Underline

OK Cancel Apply |< >| current Help

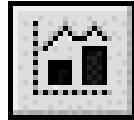
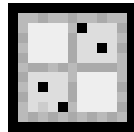
Result



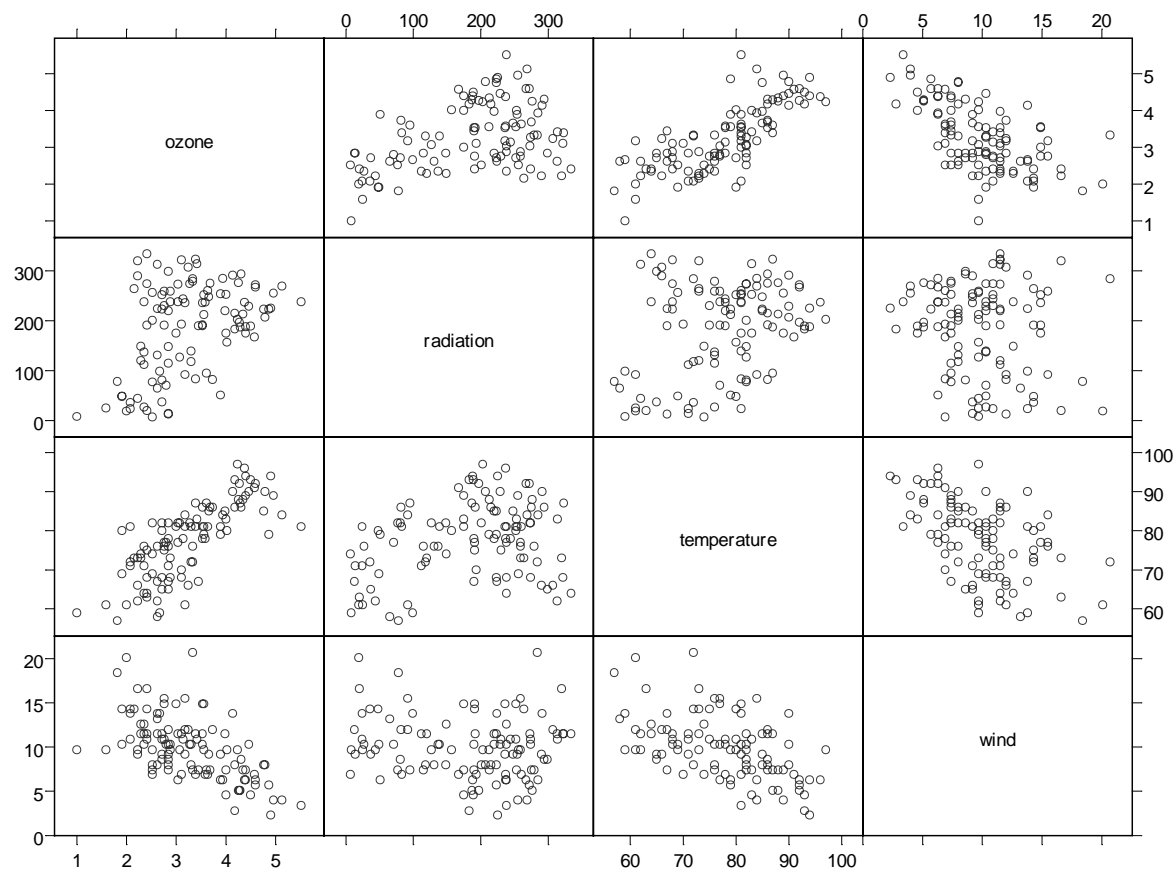
# Scatter Plot Matrix

- A scatter Plot Matrix produces a grid of scatter plots, plotting each variable against all other variables
- For the air data
  - create a scatter plot matrix
  - Add a Loess Smooth Curve

# Creating Scatter Plot Matrix

- Open the Data Window by double clicking on air in the object browser page.
- Select all of the columns
- Open the “2D plot palette”. 
- Select “Scatter Plot Matrix”. 

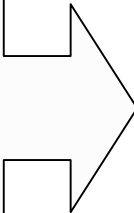
**Result**



# **Adding a Loess Smooth Curve**

- Double-click on a data point in the scatter plot
- Click on the Smooth tab
- In the Smoothing Type field, choose Loess from the pull-down menu
- Click Ok

Step 1



Scatter Plot Matrix [1]

Data to Plot | Line/Histogram | Symbol | Vary Symbols | Smooth

Data Columns

Data Set:

x Columns:

y Columns:

z Columns:

w Columns:

Override Conditioning

Type:

Data Set:

Columns:

Draw in Panels:

Data Tips and Point Labels

Column:

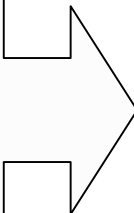
☐ Hide

☒ Crop

Subset Rows with:

OK Cancel Apply |< >| current Help

Step 2



Scatter Plot Matrix [1]

Data to Plot | Line/Histogram | Symbol | Vary Symbols | Smooth

Smoothing Type:

# Output Points:

Loess/Friedman Specs

Span:

Degree:

Family:

Smoothing Spline Specs

Deg. of Freedom:

Kernel Specs

Bandwidth:

Kernel:

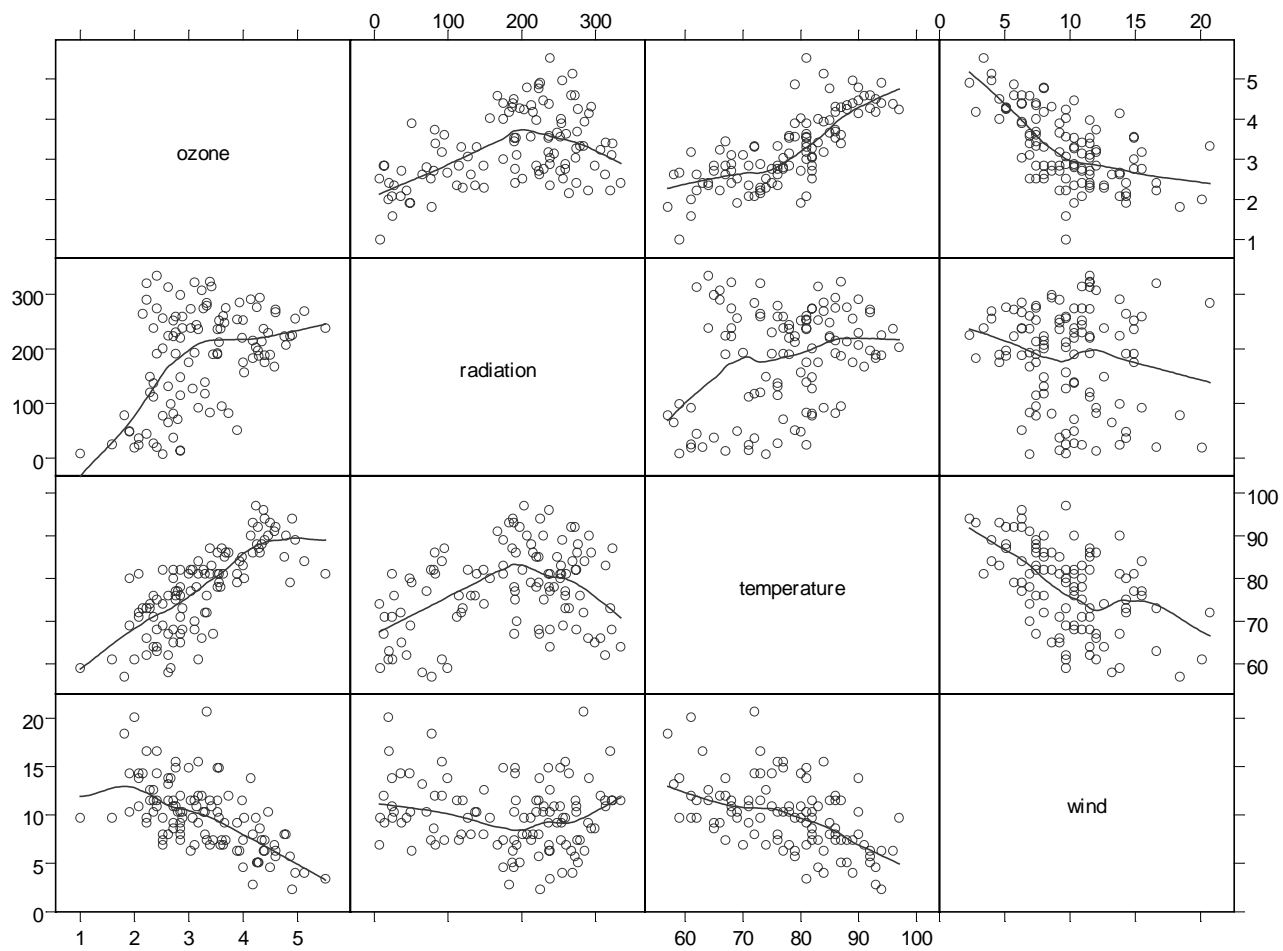
User-Defined Smoothing

Function Name:

Other Arguments:

OK Cancel Apply |< >| current Help

**Result**



# **Placing Multiple Graph on a Graph Sheet**

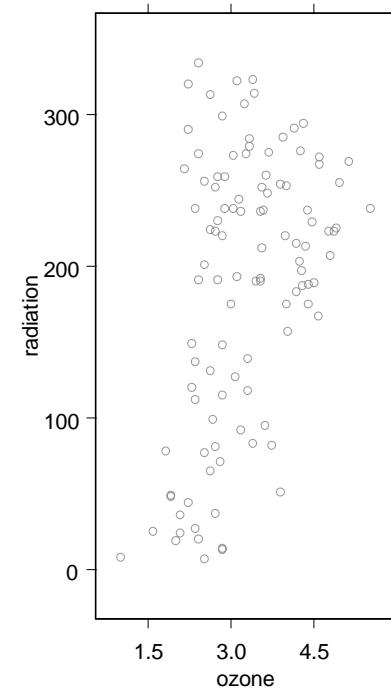
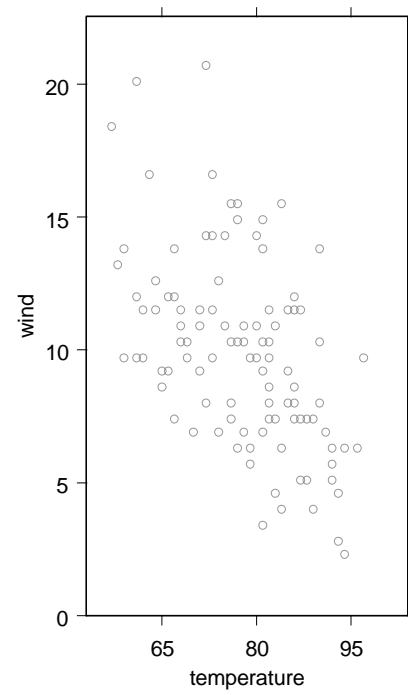
- Add a new graph to an existing graph
  - Insert/Graph from the main menu
  - Or,
    - Open the Graph sheet you wish to add a graph
    - Open the data window containing the data to plot
    - Select the Data Columns
    - Open the 2D Palette if it is not open
    - SHIFT-click the plot button on the palette.

- Combining Graphs from different graph sheets
  - Open all individual graph sheets you want to combine
  - Choose one of the following
    - Use drag/Drop
    - Open a new Graph sheet and copy all the Graph sheets you want to combine into this new one as follows
      - Click the New Button
      - Click Graph Sheet
      - Select one of the individual graph sheet
      - From the Edit Menu, choose Copy Graph Sheet Page
      - Select the empty graph sheet window
      - From the Edit menu, choose Paste
      - Repeat the above Steps for each individual graph sheet

## **Example:**

- Create a 2D scatter plot for the two variables temperature and wind.
- Insert another 2D scatter plot for the two variables ozone and radiation Using one of the above methods.
- I will be showing this in class

**Result**



# Creating Multiple Plots

- From the air data plot the following
  - **temperature** to be on the **x-axis**
  - **wind** and **ozone** on the y-axis
- To do so
  - Select the variables by selecting **temperature** first and then CTRL-click on both **wind** and **ozone**
  - Open the “2D plot palette”
  - Select “scatter plot”

**Step 1**



S-PLUS - [air]

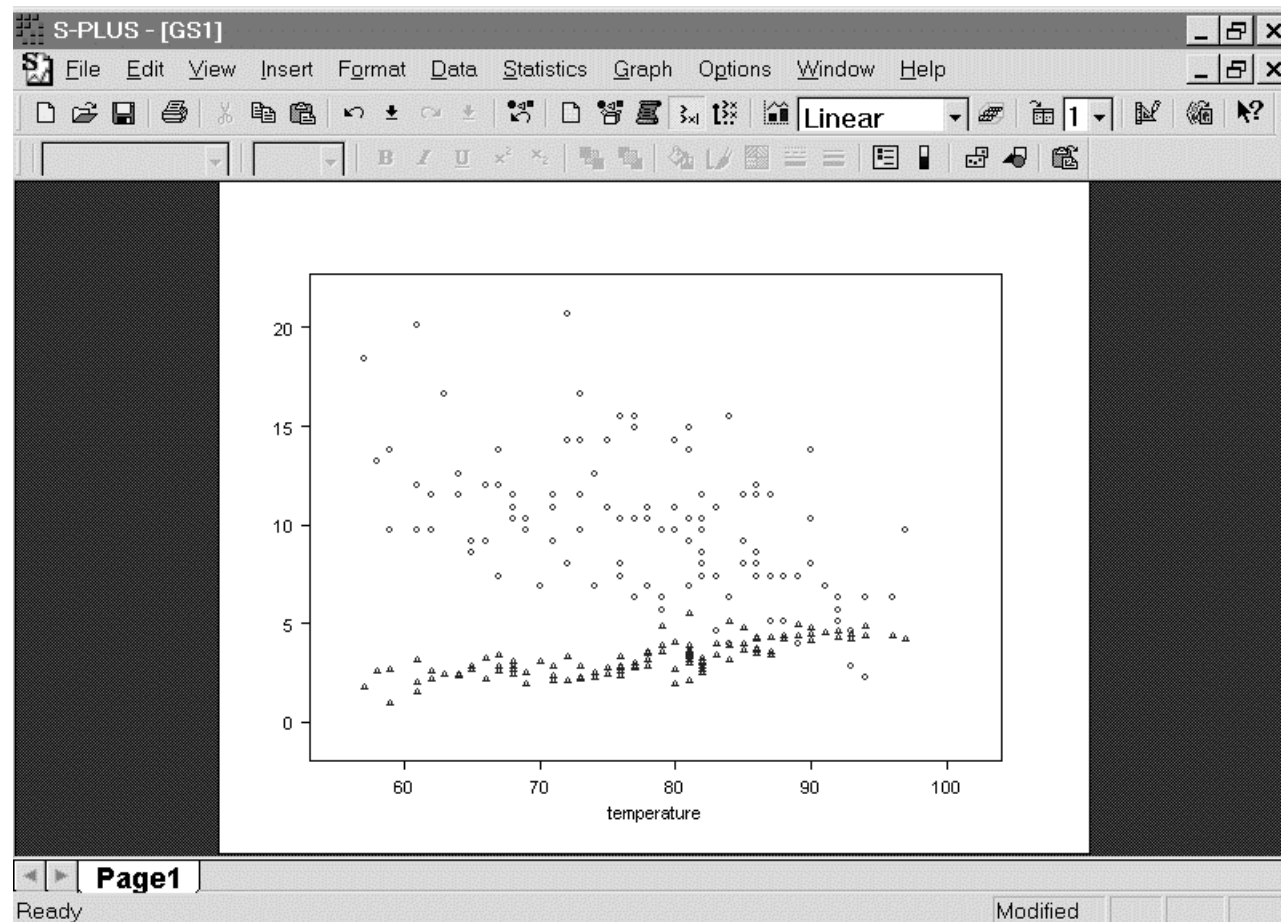
File Edit View Insert Format Data Statistics Graph Options Window Help

Linear 1

	1	2	3	4	5	6	7
	ozone	radiation	temperature	wind			
1	3.45	190.00	67.00	7.40			
2	3.30	118.00	72.00	8.00			
3	2.29	149.00	74.00	12.60			
4	2.62	313.00	62.00	11.50			
5	2.84	299.00	65.00	8.60			
6	2.67	99.00	59.00	13.80			
7	2.00	19.00	61.00	20.10			
8	2.52	256.00	69.00	9.70			
9	2.22	290.00	66.00	9.20			
10	2.41	274.00	68.00	10.90			
11	2.62	65.00	58.00	13.20			
12	2.41	334.00	64.00	11.50			
13	3.24	307.00	66.00	12.00			
14	1.82	78.00	57.00	18.40			
15	3.11	322.00	68.00	11.50			

Ready

**Result**



# 3D PLOTS

- **Example 1: 3D Spline Plot**
  - kyphosis built-in-data frame
    - Create a 3D Spline Plot for the Age, Number and Start variables.
    - Modify 3D Axis Title.
    - Rotating 3D Plots Interactively

# **Creating 3D Spline Plot**

- Open the kyphosis data set.
- Select Age, Number and Start
- Open the 3D Plot Palette
- Click on a 3D Spline plot Button

Step 1



S-PLUS - [kyphosis]

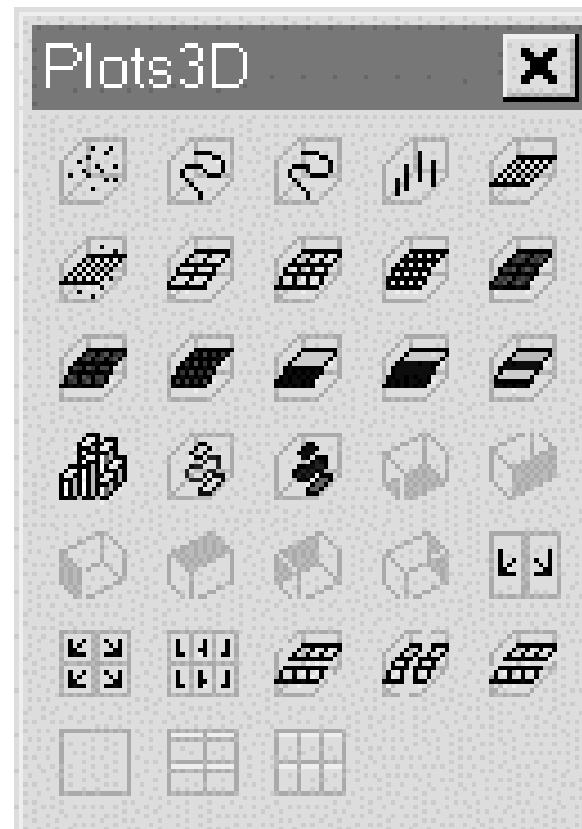
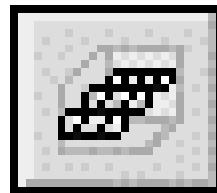
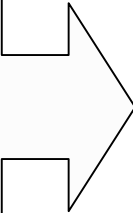
File Edit View Insert Format Data Statistics Graph Options Window Help

Linear 1

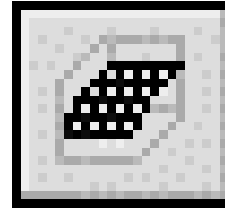
	1	2	3	4	5	6	7
	Kyphosis	Age	Number	Start			
1	absent	71.00	3.00	5.00			
2	absent	158.00	3.00	14.00			
3	present	128.00	4.00	5.00			
4	absent	2.00	5.00	1.00			
5	absent	1.00	4.00	15.00			
6	absent	1.00	2.00	16.00			
7	absent	61.00	2.00	17.00			
8	absent	37.00	3.00	16.00			
9	absent	113.00	2.00	16.00			
10	present	59.00	6.00	12.00			
11	present	82.00	5.00	14.00			
12	absent	148.00	3.00	16.00			
13	absent	18.00	5.00	2.00			
14	absent	1.00	4.00	12.00			
15	absent	168.00	3.00	18.00			

Ready

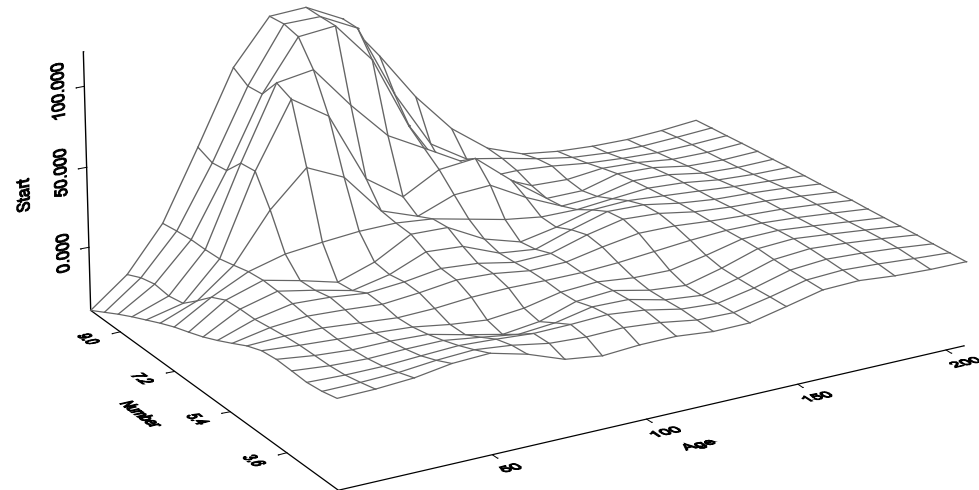
Step 2



Step 3



Result

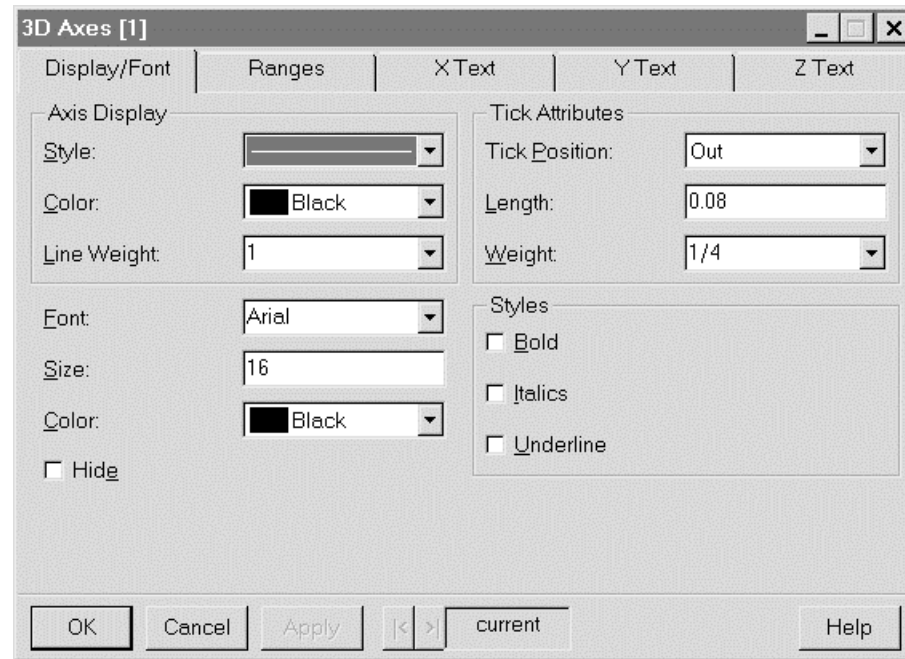


# Modifying 3D Axes Title

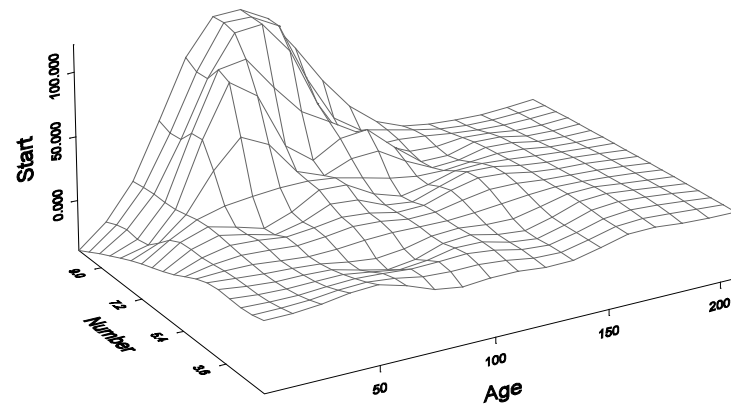
- Note:
  - 3D axis title are different from 2D axis title in that they are not separate objects from the axes. They cannot be moved and sized interactively but can be edit through a dialog
- To Modify
  - Double-click on the axes to open the 3D Axis
  - Change the font size for each axis title
    - Choose the **X Text** tab
    - In the X title section Specify **20** for the size.
    - Repeat for the Y and the Z titles

- Change the Display Plane for the X axis title
  - In the X Label Format section in the Dialog, choose **XZ** from the **Plane** Pull-down list
  - Click **Ok**

3D Axes  
Dialog

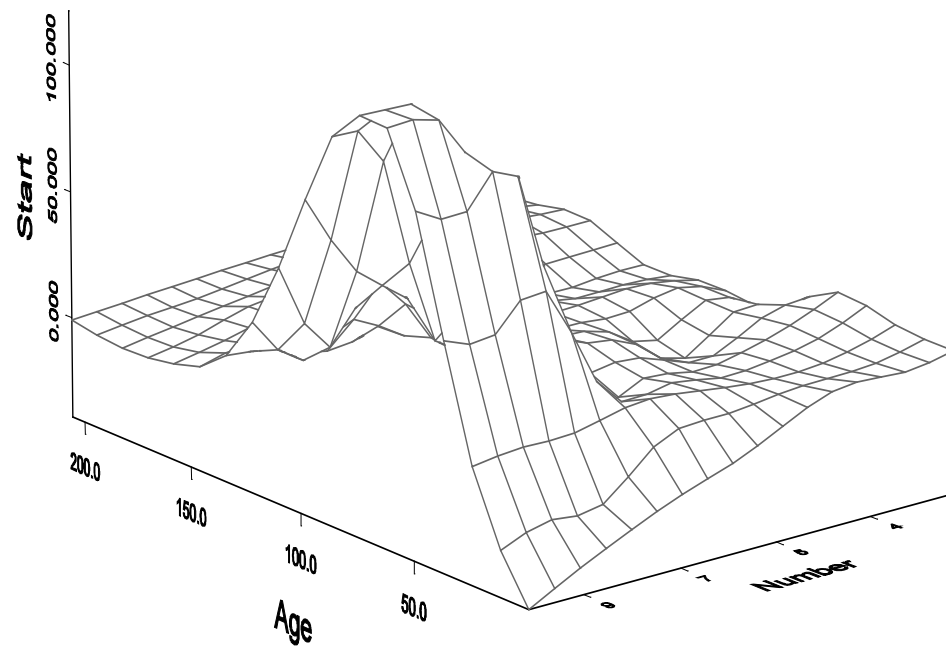
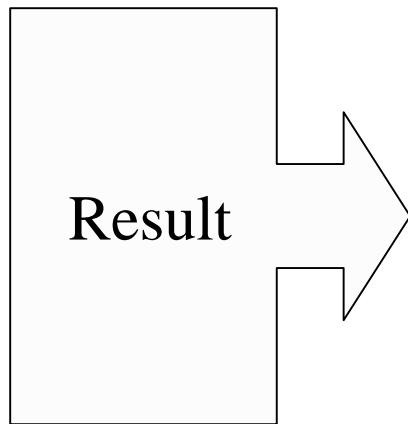


Result



# Rotating 3D Plots Interactively

- Select the 3D workbox
  - click outside the surface plot but inside the 3D workbox
  - Circular and triangular knobs appear
    - Circular knobs: rotate the workbox horizontally (around the z-axis)
    - Triangular knobs: rotate the workbox vertically (around the x-axis)
  - Rotate the workbox by dragging one of the knobs
  - Release the rotation knob.
  - The graph is redrawn at the new workbox orientation.



# Plotting Subsets of Rows

- Subsetting Rows in the Data Window
- Subsetting Rows Using an Expression
- Adding a Curve Fit Equation

## Example: Subsetting Rows in the Data Window

- For the Type=Small, create a scatter plot for Weight and Mileage variable in fuel.frame built-in-data frame
  - Open the fuel.frame data in the Data Window
  - Select the column to plot
    - CTRL-click on **Weight** and **Mileage**
    - Select the rows to plot. In this case it is the first 13 cases by holding down the CTRL key while dragging down the row header till line 13.
  - Click on the **Scatter Plot** Button on the **2D Plot palette**

Selecting  
Data

S-PLUS - [fuel.frame]

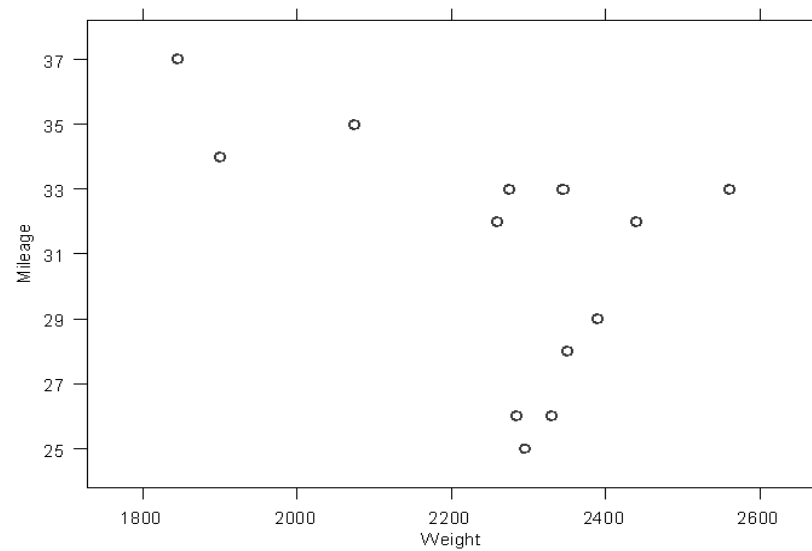
File Edit View Insert Format Data Statistics Graph Options Window Help

Linear

		1	2	3	4	5	6
		Weight	Disp.	Mileage	Fuel	Type	
1	Eagle Summit 4	2560.00	97.00	33.00	3.03	Small	
2	Ford Escort 4	2345.00	114.00	33.00	3.03	Small	
3	Ford Festiva 4	1845.00	81.00	37.00	2.70	Small	
4	Honda Civic 4	2260.00	91.00	32.00	3.13	Small	
5	Mazda Protege 4	2440.00	113.00	32.00	3.13	Small	
6	Mercury Tracer 4	2285.00	97.00	26.00	3.85	Small	
7	Nissan Sentra 4	2275.00	97.00	33.00	3.03	Small	
8	Pontiac LeMans 4	2350.00	98.00	28.00	3.57	Small	
9	Subaru Loyale 4	2295.00	109.00	25.00	4.00	Small	
10	Subaru Justy 3	1900.00	73.00	34.00	2.94	Small	
11	Toyota Corolla 4	2390.00	97.00	29.00	3.45	Small	
12	Toyota Tercel 4	2075.00	89.00	35.00	2.86	Small	
13	Volkswagen Jetta 4	2330.00	109.00	26.00	3.85	Small	
14	Chevrolet Camaro V8	3320.00	305.00	20.00	5.00	Sporty	
15	Dodge Daytona	2885.00	153.00	27.00	3.70	Sporty	

Ready

Result



## **Example: Subsetting Rows Using an Expression**

- For the Type=Small, create a scatter plot for Weight and Mileage variable in fuel.frame built-in-data frame
- Open the Dialog window by double clicking on one of the data points on the scatter plot
- Enter Expression in the **Subset Rows** plot on the **Data to Plot** page of the the **Line/Scatter Plot** dialog
  - Type= “Small”

Scatter  
Dialog

Line/Scatter Plot [1]

Data to Plot | Line | Symbol | Vary Symbols | Smooth/Sort

Data Columns

Data Set: fuel.frame

x Columns: Weight

y Columns: Mileage

z Columns:

w Columns:

Override Conditioning

Type: Auto

Data Set: fuel.frame

Columns:

Draw in Panels: All

Data Tips and Point Labels

Column: <AUTO>

Scale to

X Axis #: 1

Y Axis #: 1

Plane #: 1

☒ Use for Aspect Ratio

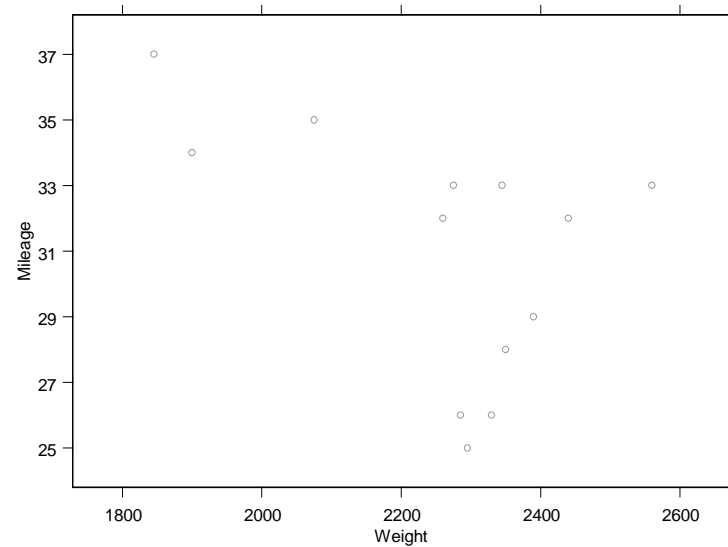
☐ Hide

☒ Crop

Subset Rows with: Type == "Small"

OK Cancel Apply < > current Help

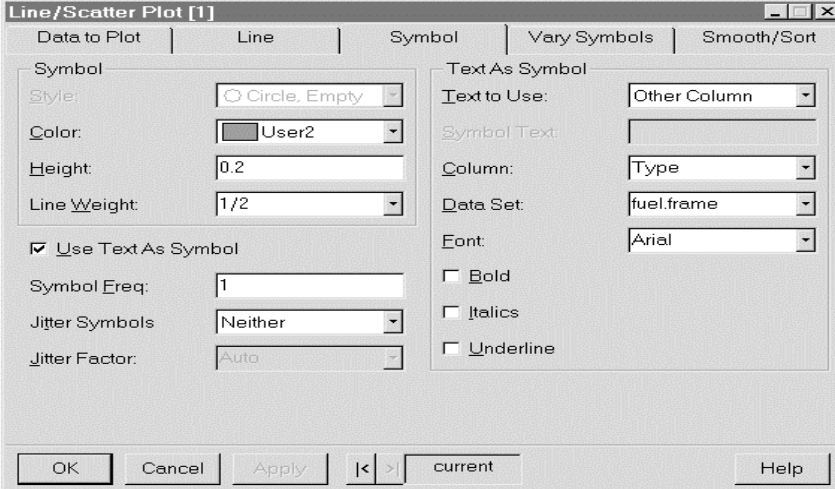
Result



# Using Text as Symbols

- Use the word “small” as the symbol in the previous plot
  - In the **Line/Scatter plot** dialog, click on **Symbol**.
  - Check the box labeled **Use Text as Symbol**.
    - In the **Text to Use** choose **Other Column**.
    - In the **Column** choose the **Type** variable.

# Line/Scatter Plot Properties Dialog



The dialog box is titled "Line/Scatter Plot [1]" and has five tabs: "Data to Plot", "Line", "Symbol", "Vary Symbols", and "Smooth/Sort". The "Symbol" tab is currently selected. It contains two main sections: "Symbol" and "Text As Symbol".

**Symbol Section:**

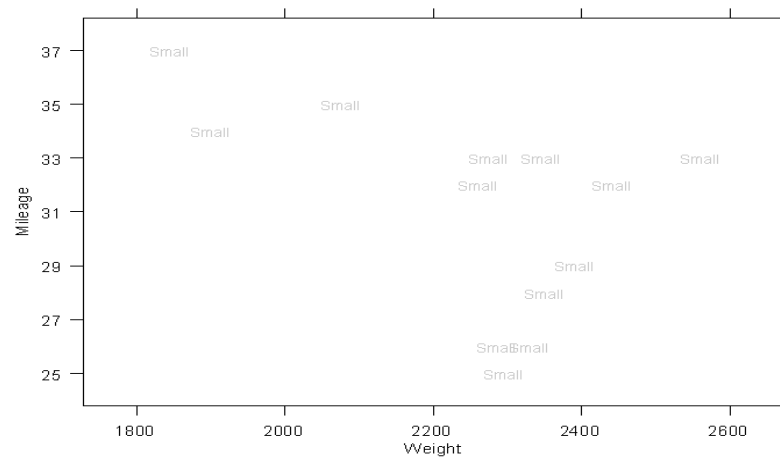
- Style: ☐ Circle, Empty
- Color: ☐ User2
- Height: 0.2
- Line Weight: 1/2
- ☒ Use Text As Symbol
- Symbol Freq: 1
- Jitter Symbols: Neither
- Jitter Factor: Auto

**Text As Symbol Section:**

- Text to Use: Other Column
- Symbol Text: (empty text box)
- Column: Type
- Data Set: fuel.frame
- Font: Arial
- ☐ Bold
- ☐ Italics
- ☐ Underline

At the bottom, there are buttons for "OK", "Cancel", "Apply", and "Help". There are also navigation buttons "<" and ">" and a "current" button.

# Result



# BRUSH and SPIN

- Brush and Spin is a powerful way to explore your data visually.
  - It can be used for the following types of plots
    - Scatter: select one point and see varying views of that point
    - Histogram: reflect highlighted and non-highlighted points for each variable
    - 3D: use a 3D point cloud projection to examine the variables
- Highlighting and Downlighting Points
- Searching for Cases
- Spinning Data

# Example: Using Brush and Spin

- Open the **fuel.frame** data set.
- Open the **Brush and Spin** dialog
  - Click **Graph**
  - Click **Brush and Spin**
- Select the columns to plot
  - In the column pull-down list, CTRL-click to select all of the columns in **fuel.frame**
- Click **Ok**
- To quit click QUIT in the Brush window

Step 1

S-PLUS - [fuel.frame]

		1	2	3	4	5
		Weight	Disp.	Mileage	Fuel	Type
1	Eagle Summit 4	2560.00	97.00	33.00	3.03	Small
2	Ford Escort 4	2345.00	114.00	33.00	3.03	Small
3	Ford Festiva 4	1845.00	81.00	37.00	2.70	Small
4	Honda Civic 4	2260.00	91.00	32.00	3.13	Small
5	Mazda Protege 4	2440.00	113.00	32.00	3.13	Small
6	Mercury Tracer 4	2285.00	97.00	26.00	3.85	Small
7	Nissan Sentra 4	2275.00	97.00	33.00	3.03	Small
8	Pontiac LeMans 4	2350.00	98.00	28.00	3.57	Small
9	Subaru Loyale 4	2295.00	109.00	25.00	4.00	Small
10	Subaru Justy 3	1900.00	73.00	34.00	2.94	Small
11	Toyota Corolla 4	2390.00	97.00	29.00	3.45	Small
12	Toyota Tercel 4	2075.00	89.00	35.00	2.86	Small
13	Volkswagen Jetta 4	2330.00	109.00	26.00	3.85	Small
14	Chevrolet Camaro V8	3320.00	305.00	20.00	5.00	Sporty

Ready NUM

Step 2

Brush Properties [1]

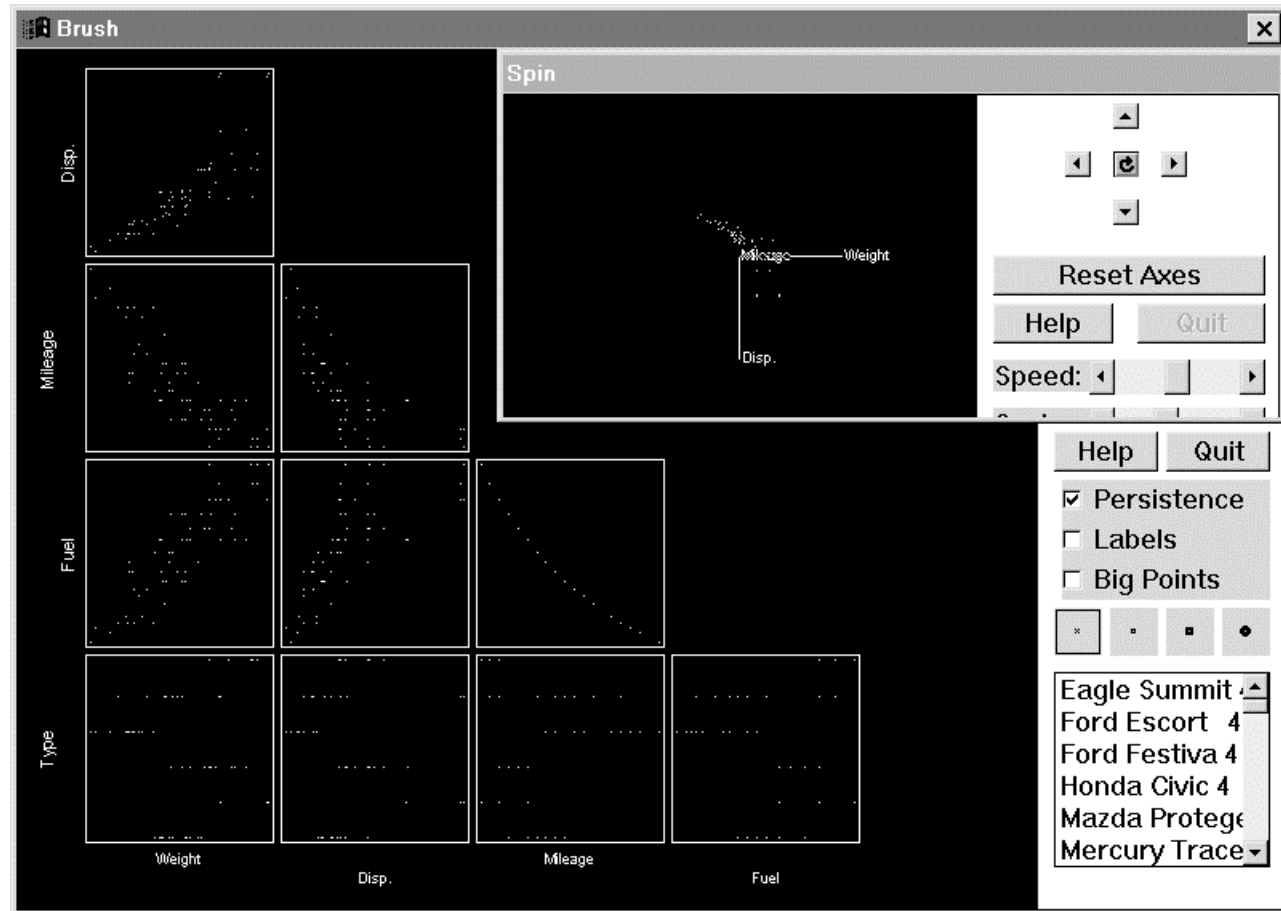
Brush and Spin

Data Set: fuel.frame ☐ Draw Histograms

Columns: ALL ☒ Spin Window

OK Cancel |< >| current Help

Result



- The Brush Window displays a scatter plot matrix of the variables
- The spin window shows a 3D representation of the first 3 selected variables

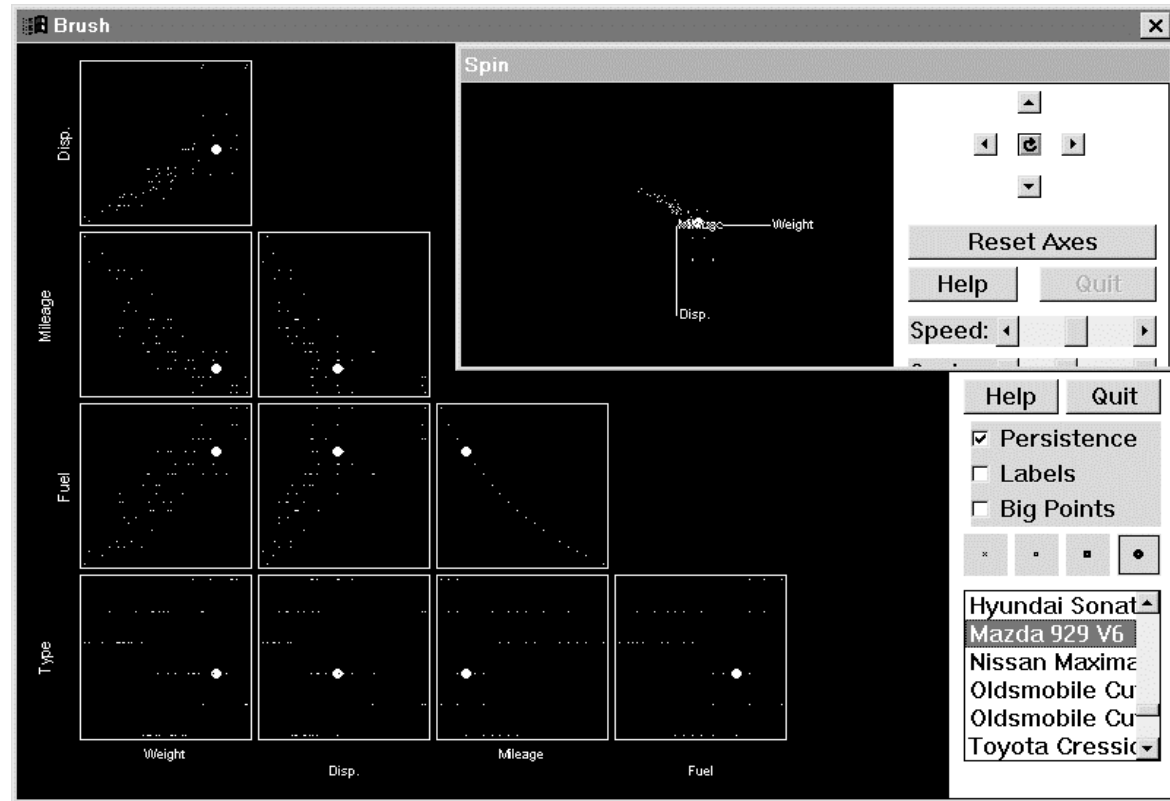
# **Highlighting and Downlighting Points**

- In the Brush window, the mouse cursor appears with a box, called the brush, when it passes inside any of the scatter plots.
- To highlight some of the points
  - Move the mouse cursor into one of the scatter plots so the brush covers some data points
  - Click the left mouse button
- To Downlight some of the points
  - Click the right mouse button

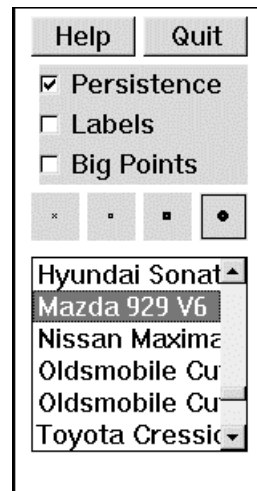
# Searching for Cases

- To search for cases we use the list box that contains the row names or numbers.
- Suppose we are searching for **Masda 929 V6**.
  - Search for it in the list box
  - Left click on the name
  - Increase the highlight symbol size

Result

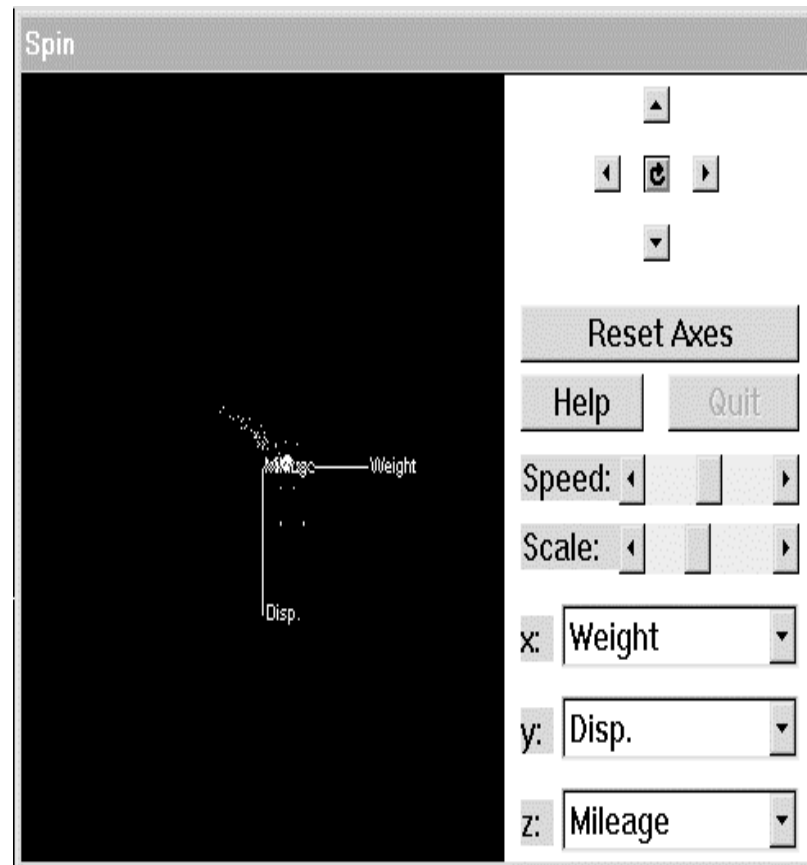


Brush  
Window



# Spinning Data

- Use the spin Control to rotate the 3D plot
- Use speed control to adjust the speed of the spinning
- Use x, y and z prompts to change the variables being plotted



# Basic Statistical Models in S-PLUS

# Objectives:

- Introduce some of the S-Plus statistical modeling dialogs like
  - Simple Summary Statistics
  - Simple Linear regression

# Contents:

- Statistics Dialogs
- Dialog Control and Fields
- Computing Summary Statistics
- Linear regression
- Data Transformations
- Output

# Statistics Dialogs

- Much of the Statistical functionality of S-Plus can be accessed through the **Data** and **Statistics** Options on the main menu.

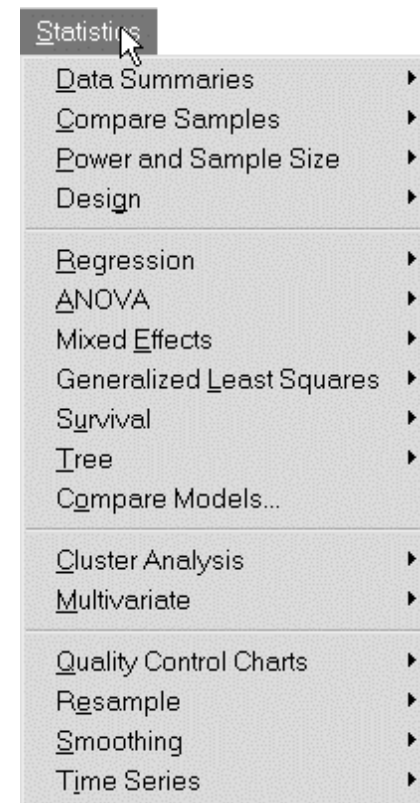
# Data Menu

- It includes dialogs for the following:
  - Tabulating data
  - Calculate Distribution functions
  - Generating random samples
  - Generating random numbers



# Statistics Menu

- It includes dialogs for the following:
  - Creating data summaries
  - Performing hypothesis tests
  - Fitting Statistical Models



# Dialog Controls

- There are several common controls which execute each dialog

- Rollback arrows

- These arrows move through previous states of the dialog



- Apply button: When clicking on this button

- You accept the current state of the dialog
  - Execute the corresponding function
  - Keep the dialog open



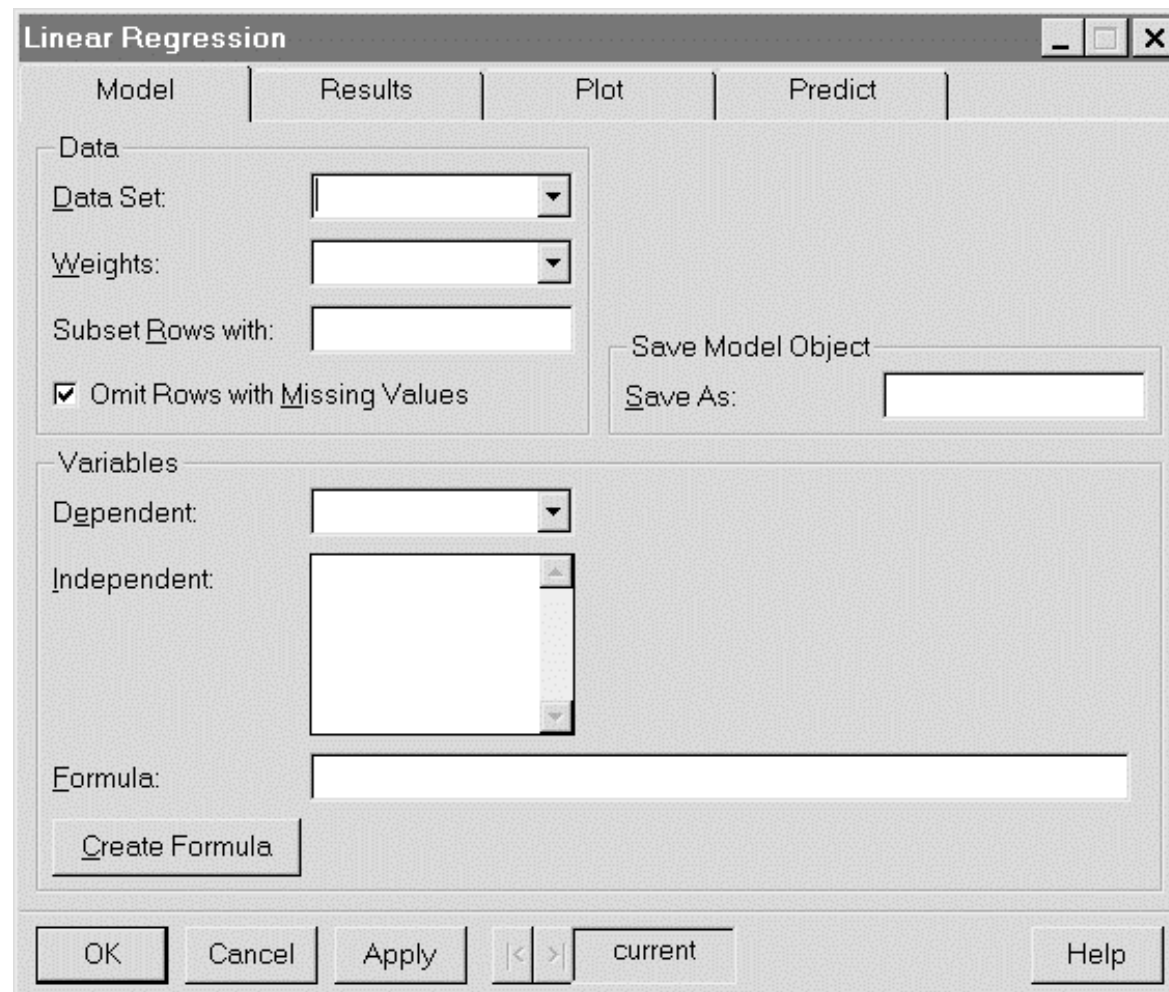
- Ok button: When clicking on this button
  - You accept the current state of the dialog
  - Execute the corresponding function
  - Close the dialog



# Dialog Fields

- Each field corresponds to an argument in an S-Plus function. For example,
  - Data Frame field
    - the listed data frames are limited to those that have been filtered by an object browser
  - Formula field
    - a formula is required in Statistical modeling
  - Defaulted argument fields that might be changed
  - Optional arguments fields that generally contains appropriate defaults

# Example:



The image shows a 'Linear Regression' dialog box with a tabbed interface. The 'Model' tab is selected. The 'Data' section contains a 'Data Set' dropdown, a 'Weights' dropdown, a 'Subset Rows with:' text box, and a checked checkbox for 'Omit Rows with Missing Values'. To the right is a 'Save Model Object' section with a 'Save As:' text box. The 'Variables' section has a 'Dependent:' dropdown and an 'Independent:' list box. Below these is a 'Formula:' text box and a 'Create Formula' button. The bottom of the dialog features 'OK', 'Cancel', 'Apply', and navigation buttons ('<', '>', 'current'), along with a 'Help' button.

**Linear Regression**

Model | Results | Plot | Predict

Data

Data Set:

Weights:

Subset Rows with:

☒ Omit Rows with Missing Values

Save Model Object

Save As:

Variables

Dependent:

Independent:

Formula:

Create Formula

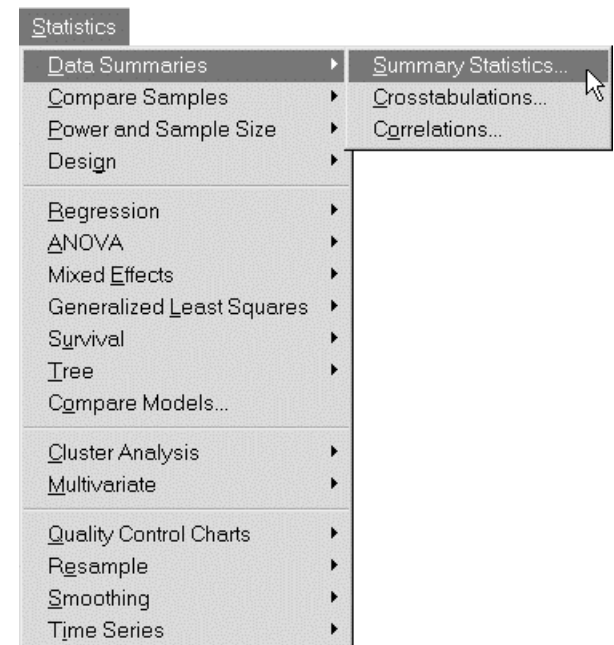
OK Cancel Apply < > current Help

# Computing Summary Statistics

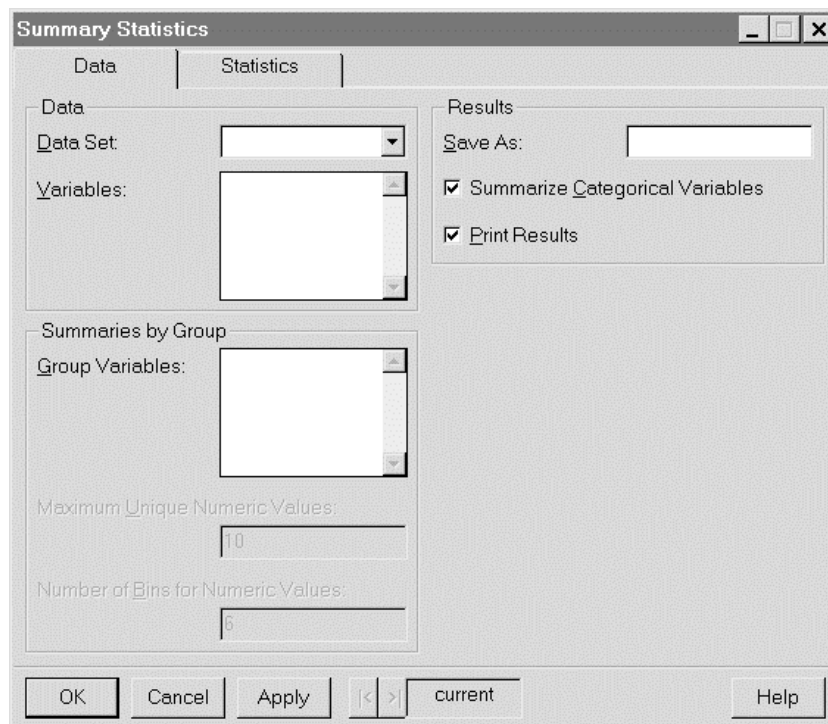
- The Summary Statistics dialogue provides basic summary statistics for a data frame, matrix or a vector of data.
- Data are summarized in tabular form in the report window
- Results can be printed and/or saved as a file.

# To Compute Summary Statistics:

- Select “Statistics” from the main menu
- Select “Data Summaries”
- Select “Summary Statistics”
- A Summary Statistics Dialog will appear
- Select the summary statistics you want to be computed for each column in the data object by checking the appropriate boxes.
- Click Ok



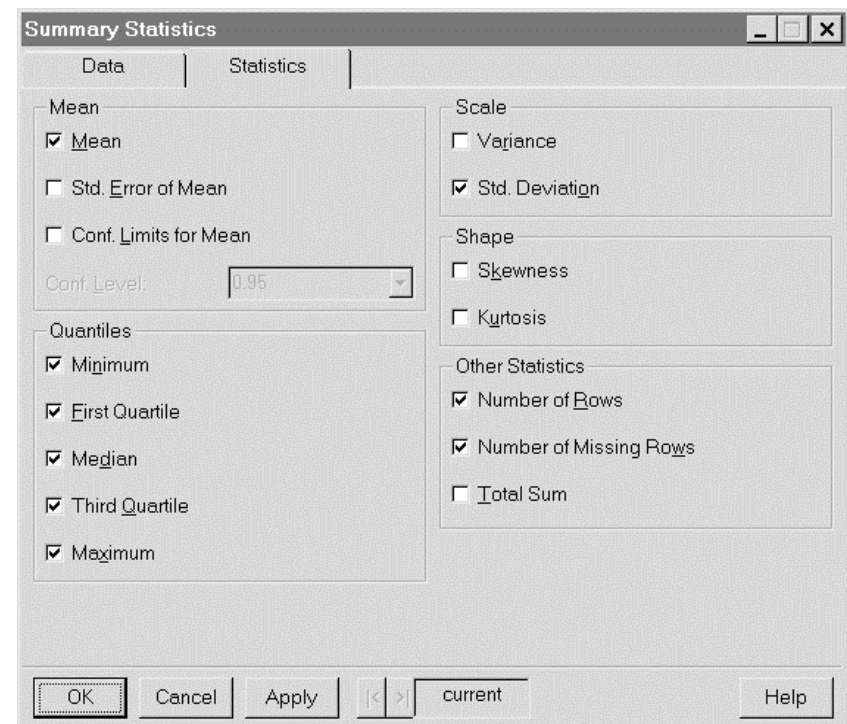
# Statistics Summary Dialog



The dialog box is titled "Summary Statistics" and has two tabs: "Data" and "Statistics". The "Data" tab is selected. It contains the following fields and options:

- Data Set:** A dropdown menu.
- Variables:** A list box.
- Summaries by Group:** A section with a "Group Variables:" list box.
- Maximum Unique Numeric Values:** A text box containing the value "10".
- Number of Bins for Numeric Values:** A text box containing the value "6".
- Results:** A section with a "Save As:" text box and two checked checkboxes: "Summarize Categorical Variables" and "Print Results".

At the bottom, there are buttons for "OK", "Cancel", "Apply", a navigation bar with "<" and ">" arrows and a "current" label, and a "Help" button.



The dialog box is titled "Summary Statistics" and has two tabs: "Data" and "Statistics". The "Statistics" tab is selected. It contains the following fields and options:

- Mean:** A section with three checkboxes: "Mean" (checked), "Std. Error of Mean", and "Conf. Limits for Mean". Below these is a "Conf. Level:" dropdown menu set to "0.95".
- Quantiles:** A section with five checked checkboxes: "Minimum", "First Quartile", "Median", "Third Quartile", and "Maximum".
- Scale:** A section with two checkboxes: "Variance" and "Std. Deviation" (checked).
- Shape:** A section with two checkboxes: "Skewness" and "Kurtosis".
- Other Statistics:** A section with three checkboxes: "Number of Rows" (checked), "Number of Missing Rows" (checked), and "Total Sum".

At the bottom, there are buttons for "OK", "Cancel", "Apply", a navigation bar with "<" and ">" arrows and a "current" label, and a "Help" button.

- **Data**
  - Data Frame
    - Select a data frame.
  - Variables
    - Select the columns of the data frame for which summary statistics will be generated.
    - To generate summary statistics for all columns, select (All Variables). Making no selection has the same effect.
  - Note: You can type into the Data Frame edit box the name of another S-PLUS object such as a vector or a matrix. In the case of a matrix, the summary statistics will be computed for the columns.

- **Summaries by Group**

- Grouping Variables

- Select the grouping variable you want to do the summary by.

- Maximum Unique Numeric Values

- Determines if the numeric grouping variables will be binned or not. If the numeric grouping variable has more distinct values than the number specified here, the variable will be binned. If not, one group will be formed for each distinct value.

- Number of Bins for Numeric Values

- Defines the number of equal length bins, the grouping numeric variable(s) will split into. If the grouping numeric variable is not going to be binned, then this number is not used.

- **Summary Statistics**
  - Minimum
    - Check this to generate the minimum value for each numeric column of the data frame.
  - First Quartile
    - Check this to generate the first quartile value for each numeric column of the data frame.
  - Mean
    - Check this to generate the mean value for each numeric column of the data frame.
  - Median
    - Check this to generate the median value for each numeric column of the data frame.

## – Third Quartile

- Check this to generate the third quartile value for each numeric column of the data frame.

## – Maximum

- Check this to generate the maximum value for each numeric column of the data frame.

## – Number of Rows

- Check this to generate the number of rows value for each numeric column of the data frame.

## – Number of Missing Rows

- Check this to generate the number of missing values (NAs) in each numeric column of the data frame.

- Variance
  - Check this to generate the variance estimate for each numeric column of the data frame.
- Std. Deviation
  - Check this to generate the standard deviation value for each numeric column of the data frame.
- Total Sum
  - Check this to generate the sum of all numeric values in each column of the data frame.
- Skewness
  - Check this to generate the skewness of all numeric values in each column of the data frame.
- Kurtosis
  - Check this to generate the kurtosis of all numeric values in each column of the data frame.

- **Results**

- Summarize Categorical Variables

- Check this to include summaries of the categorical variables (factors) in the data frame. The corresponding summaries will be
      - the factor levels
      - a count of how many values in each level are in the factor column.

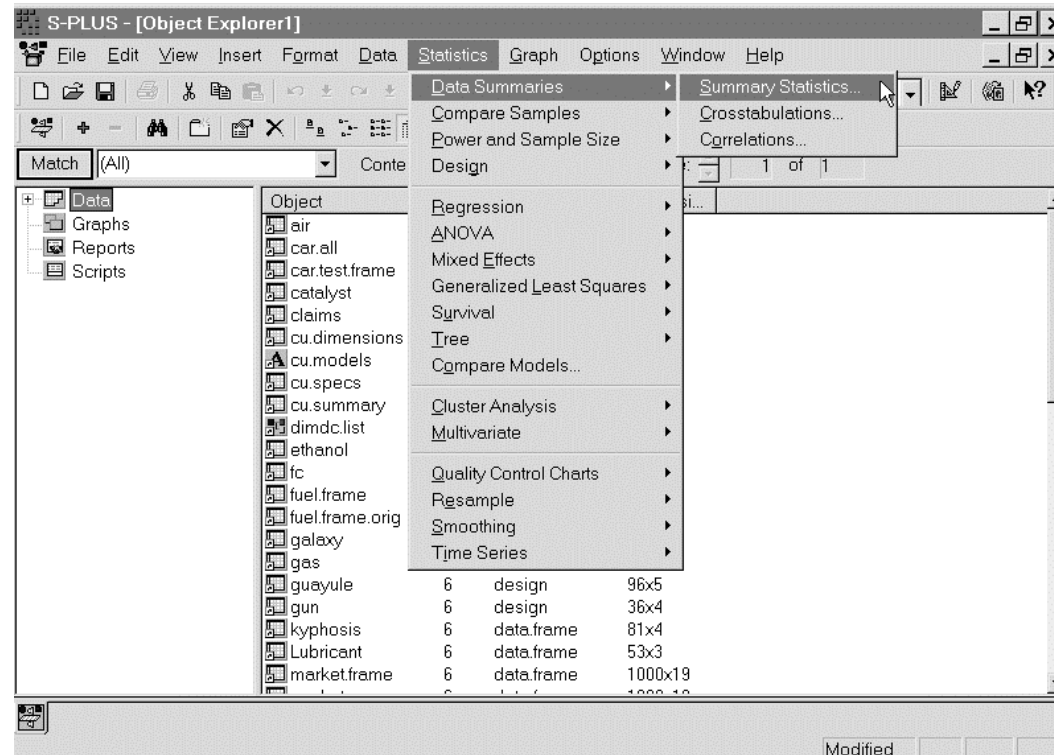
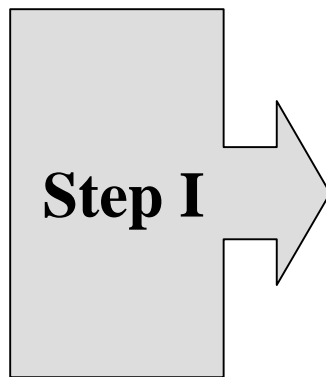
- Print Results

- Check here to display the table of counts in the designated output window.

- Save As
  - Enter the name for the object in which to save the results of the analysis. If an object with this name already exists, its contents will be overwritten.
    - This must be a valid S-PLUS object name—any combination of alphanumeric characters that starts with an alpha character is allowed. The only non-alphanumeric character allowed is the period ".". Names are case-sensitive, so X and x are different names.
- Related S-PLUS language functions:
  - min, median, mean, summary, var, max, print.table

# Example 1:

Calculate the summary statistics for the all of the variables in the “**gun**” built-in-data frame in S-Plus.



Step 2A

**Summary Statistics**

Data | Statistics

Data Set: gun

Variables: <ALL>  
Method  
Physique  
Team  
Rounds

Summaries by Group

Group Variables: (None)  
Method  
Physique  
Team  
Rounds

Maximum Unique Numeric Values: 10

Number of Bins for Numeric Values: 5

Results

Save As:

☒ Summarize Categorical Variables

☒ Print Results

OK Cancel Apply < > current Help

Step 2B

**Summary Statistics**

Data | Statistics

Mean

☒ Mean

☐ Std. Error of Mean

☐ Conf. Limits for Mean

Conf. Level: 0.95

Quantiles

☒ Minimum

☒ First Quartile

☒ Median

☒ Third Quartile

☒ Maximum

Scale

☒ Variance

☒ Std. Deviation

Shape

☒ Skewness

☒ Kurtosis

Other Statistics

☒ Number of Rows

☒ Number of Missing Rows

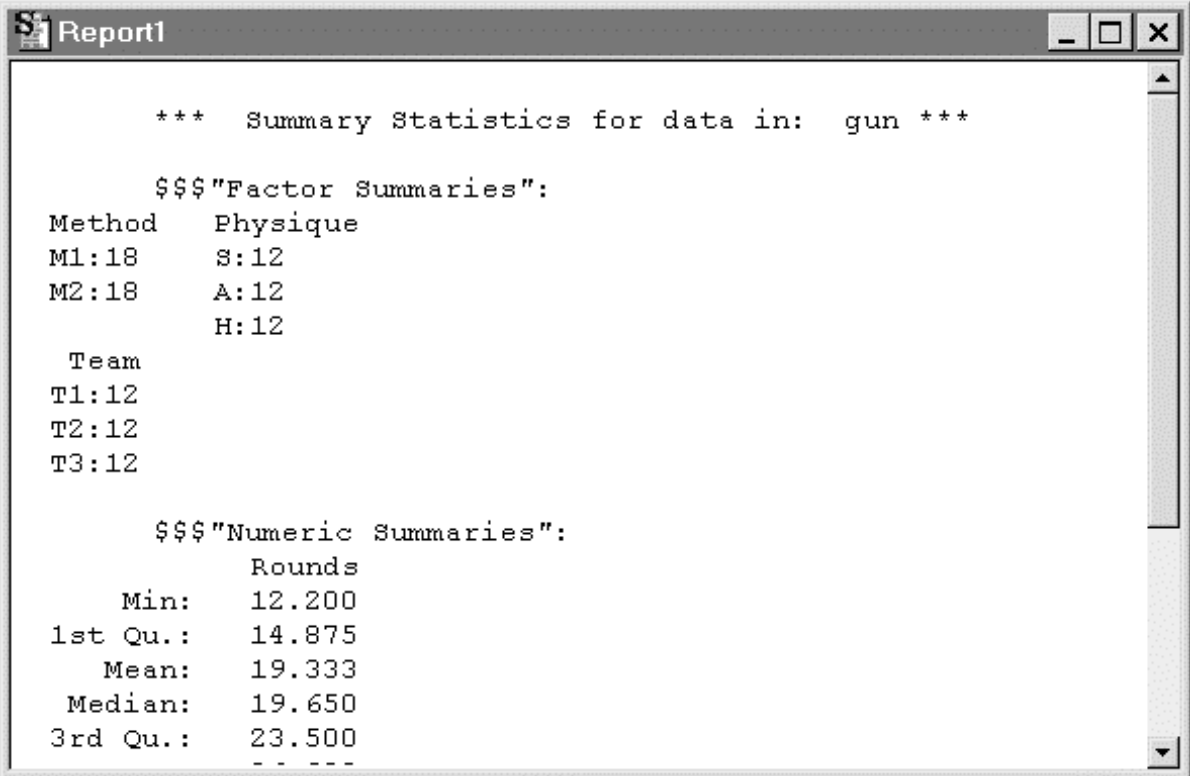
☒ Total Sum

OK Cancel Apply < > current Help

## Results



Results are printed in the Report window and also saved in the gun.summary Object which can be retrieved anytime.



```
*** Summary Statistics for data in: gun ***

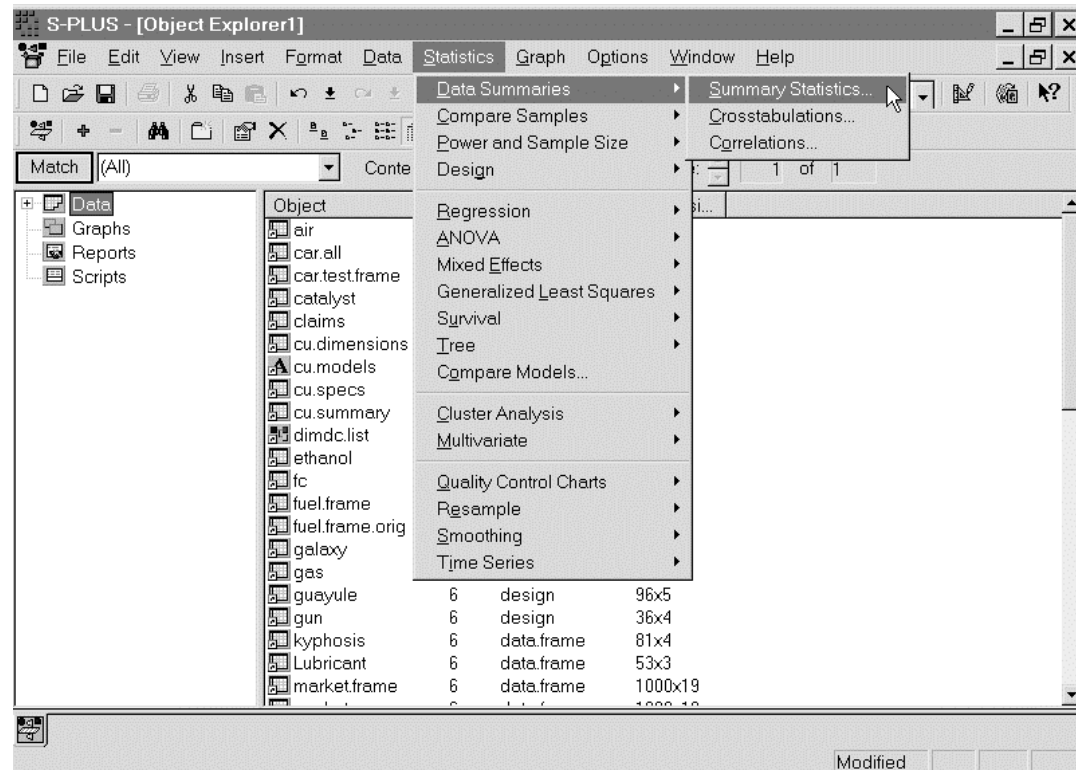
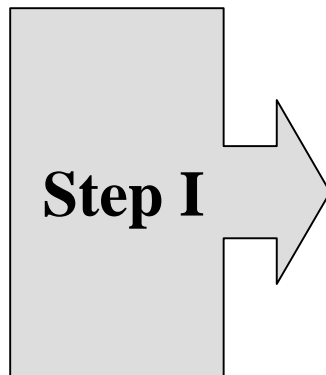
$$$"Factor Summaries":
Method    Physique
M1:18     S:12
M2:18     A:12
          H:12

Team
T1:12
T2:12
T3:12

$$$"Numeric Summaries":
          Rounds
Min:     12.200
1st Qu.: 14.875
Mean:    19.333
Median:  19.650
3rd Qu.: 23.500
-- -- --
```

# Example 2:

Calculate the summary statistics for the **Rounds** variable for each of the levels of the methods for loading rounds into Naval gun i.e. by the **Method** variable in the "gun" built-in-data frame in S-Plus.



## Step 2A

**Summary Statistics**

**Data** | **Statistics**

Data Set: gun

Variables: <ALL> Method Physique Team Rounds

Summaries by Group

Group Variables: (None) Method Physique Team Rounds

Maximum Unique Numeric Values: 10

Number of Bins for Numeric Values: 6

Results

Save As:

☒ Summarize Categorical Variables

☒ Print Results

OK Cancel Apply < > current Help

## Step 2B

**Summary Statistics**

**Data** | **Statistics**

Mean

☒ Mean

☐ Std. Error of Mean

☐ Conf. Limits for Mean

Conf. Level: 0.95

Scale

☐ Variance

☒ Std. Deviation

Shape

☐ Skewness

☐ Kurtosis

Quantiles

☒ Minimum

☒ First Quartile

☒ Median

☒ Third Quartile

☒ Maximum

Other Statistics

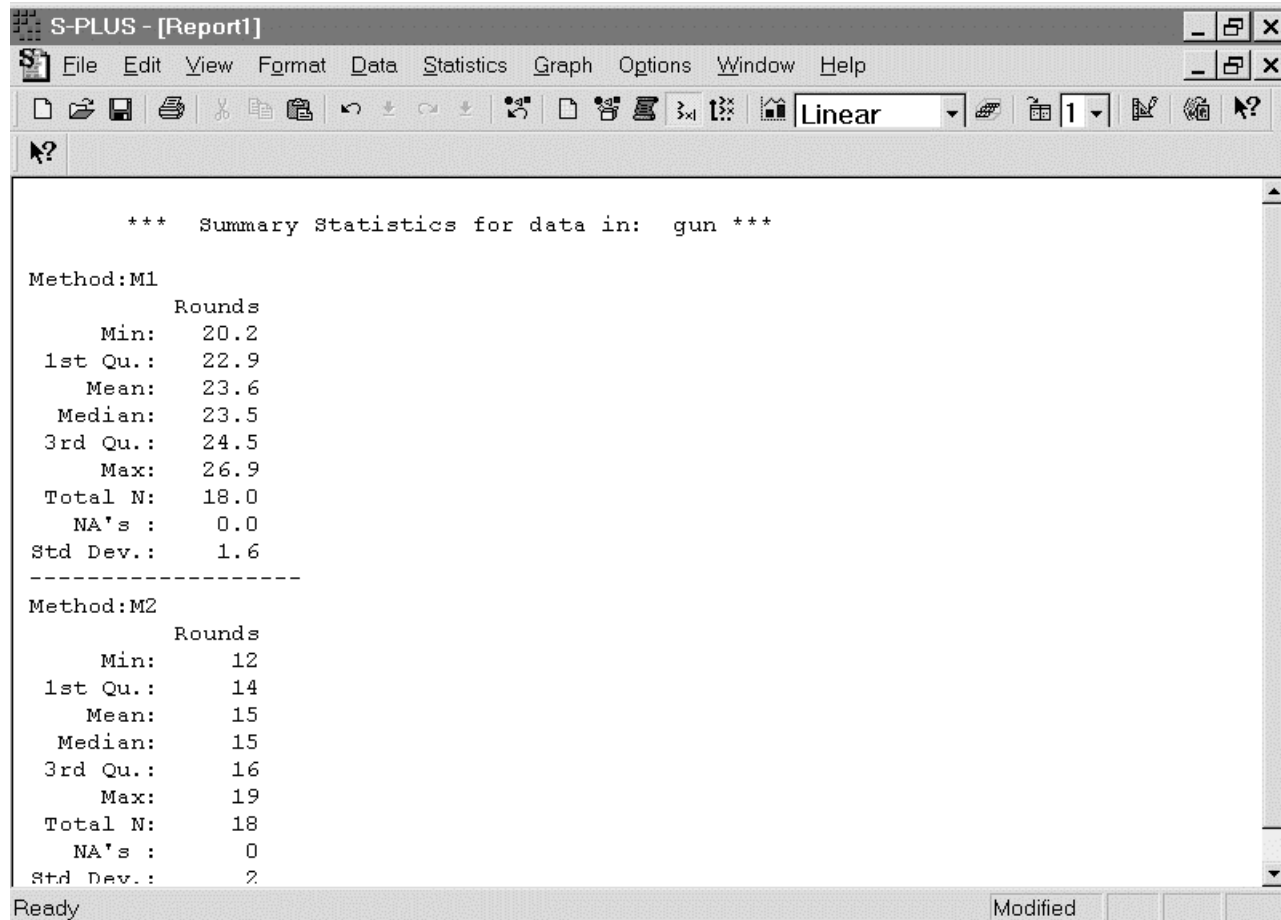
☒ Number of Rows

☒ Number of Missing Rows

☐ Total Sum

OK Cancel Apply < > current Help

# Results



# Linear Regression

- Regression Analysis includes the following:
  - Plotting the Data
  - Constructing the Model
  - Evaluating the model by Looking at the Residual.

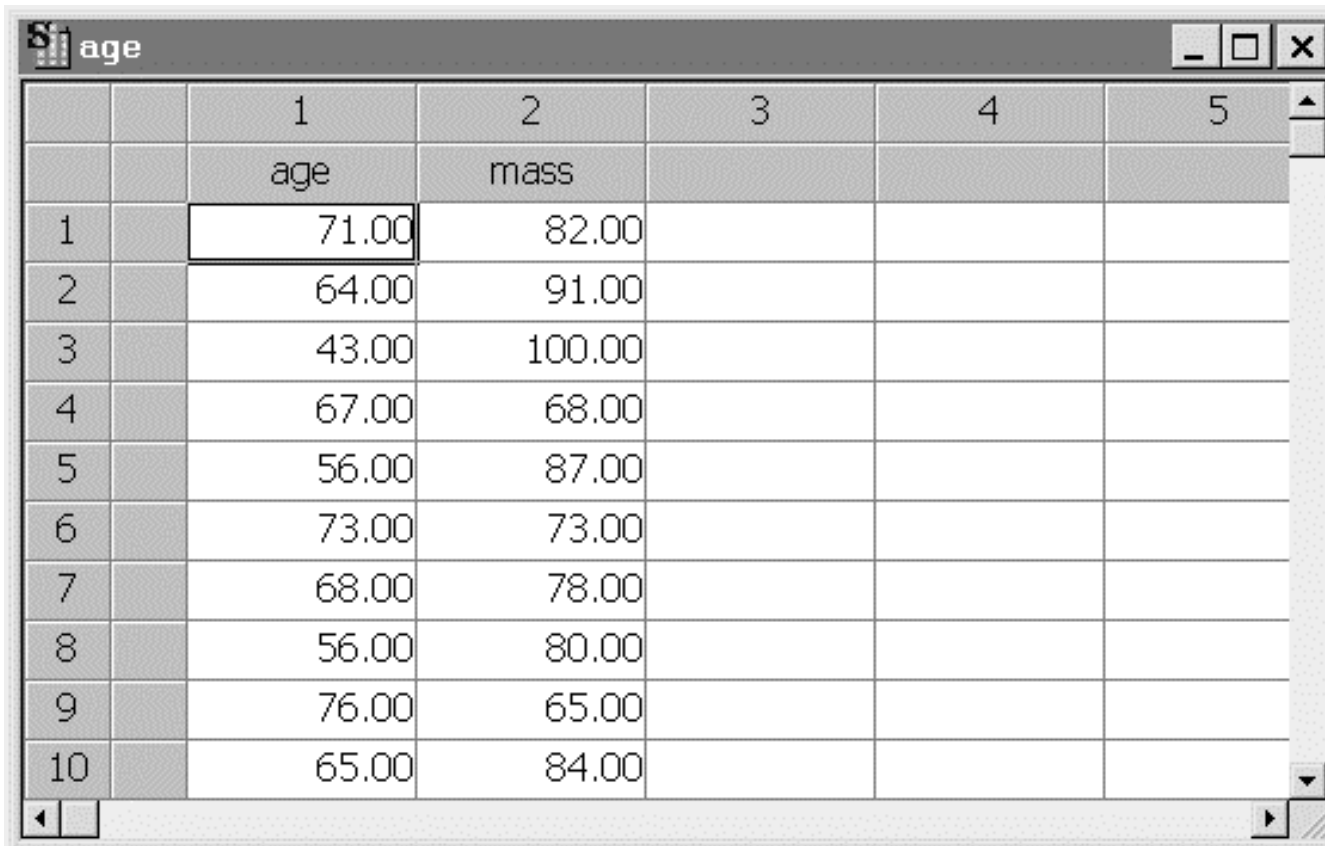
# Example:

A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected four women from each 10-year age group, beginning with age 40 and ending with age 79. The results are as follows and are saved in age.xls file:

<u>age</u>	<u>mass</u>
71	82
64	91
43	100
67	68
56	87
73	73
68	78
56	80
76	65
65	84
45	116
58	76
45	97
53	100
49	105
78	77

# Importing the Data

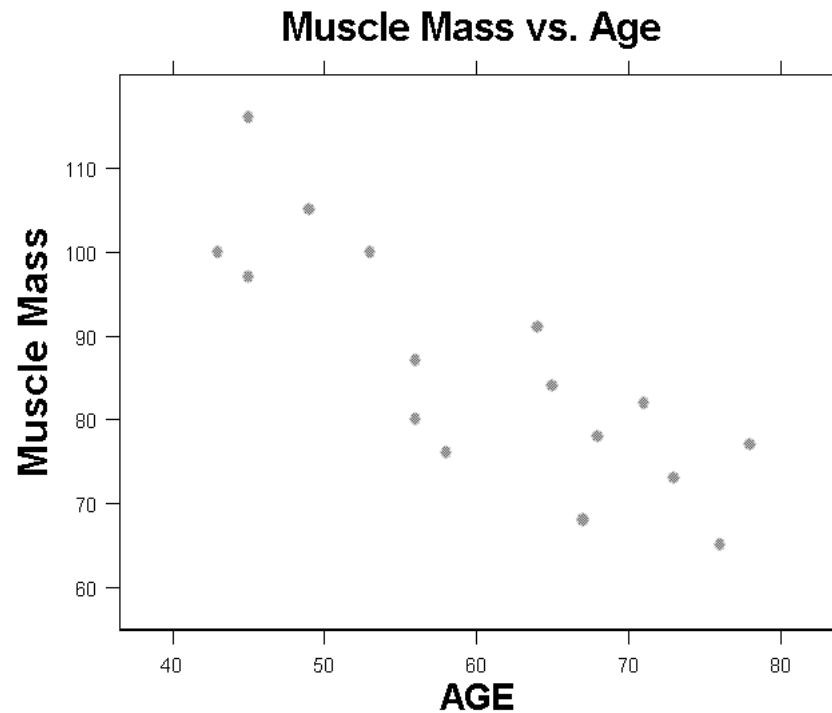
- Import the data file age.xls as we learned before



		1	2	3	4	5
		age	mass			
1		71.00	82.00			
2		64.00	91.00			
3		43.00	100.00			
4		67.00	68.00			
5		56.00	87.00			
6		73.00	73.00			
7		68.00	78.00			
8		56.00	80.00			
9		76.00	65.00			
10		65.00	84.00			

# Plotting the Data

- To Create the plot
  - Select the data columns
  - Open the 2D Palette
  - Click the 2D Scatter plot



# Constructing A Model

From the Plot we notice that a linear fit is appropriate.  
So the model that we will consider is

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

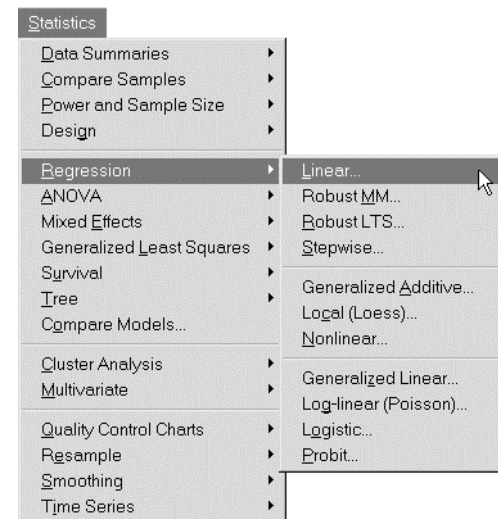
where

$Y$  = Muscle Mass

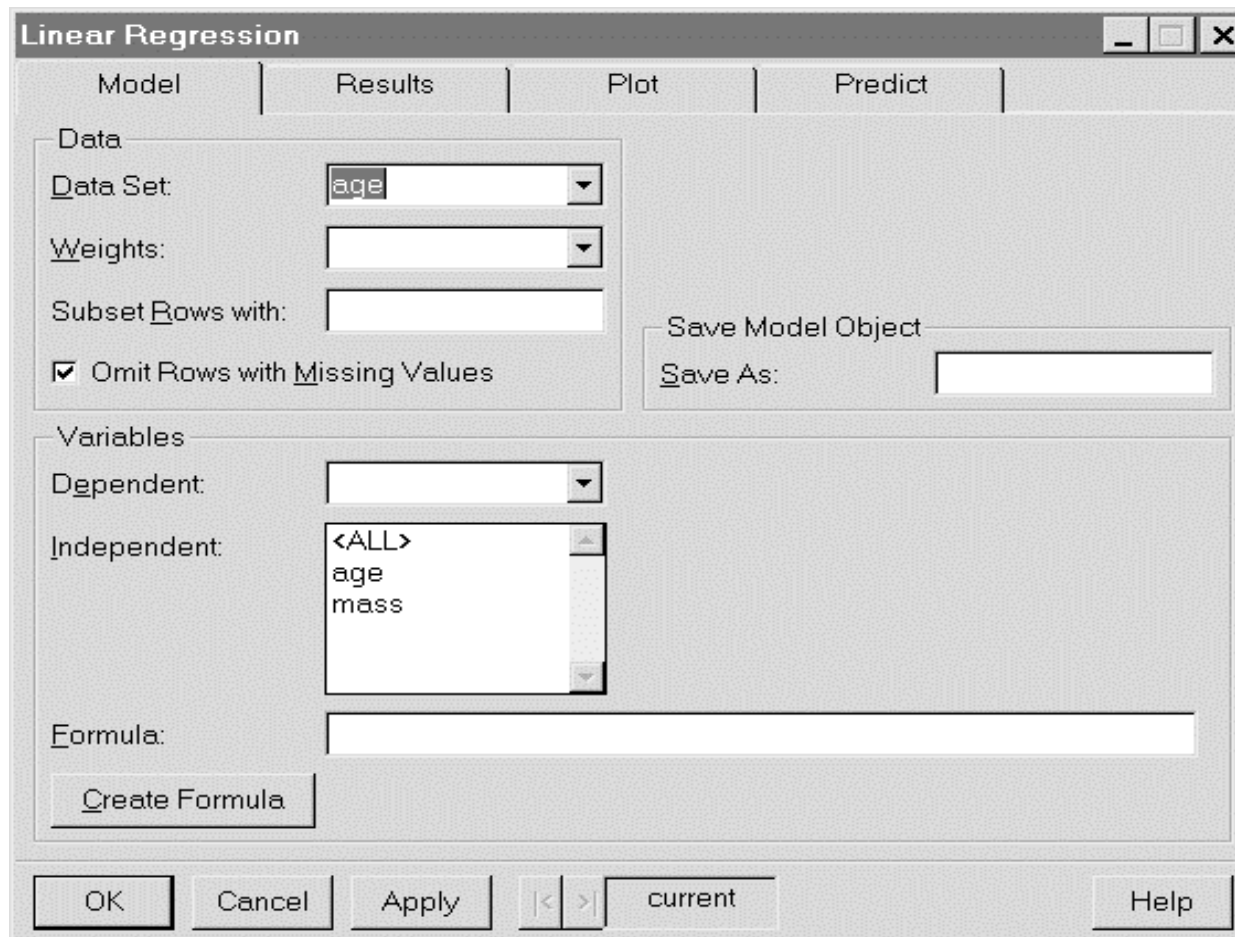
$X$  = Age

# Fitting the Model

- Select “Statistics” from the main menu
- Select “Regression”
- Select “Linear”
- A Regression Dialog will appear
- Fill in the necessary fields
- Click Ok



# Linear Regression Dialog : *Model*



The image shows a software dialog box titled "Linear Regression" with a tabbed interface. The "Model" tab is selected. The dialog is divided into several sections: "Data", "Variables", and "Formula". In the "Data" section, "Data Set" is set to "age", "Weights" is empty, "Subset Rows with:" is empty, and the checkbox "Omit Rows with Missing Values" is checked. In the "Variables" section, "Dependent:" is empty, and "Independent:" contains a list with "<ALL>", "age", and "mass". The "Formula:" section has an empty text box and a "Create Formula" button. To the right of the "Data" section is a "Save Model Object" section with a "Save As:" label and an empty text box. At the bottom are buttons for "OK", "Cancel", "Apply", navigation arrows, a "current" label, and a "Help" button.

**Linear Regression**

Model | Results | Plot | Predict

Data

Data Set: age

Weights:

Subset Rows with:

☒ Omit Rows with Missing Values

Save Model Object

Save As:

Variables

Dependent:

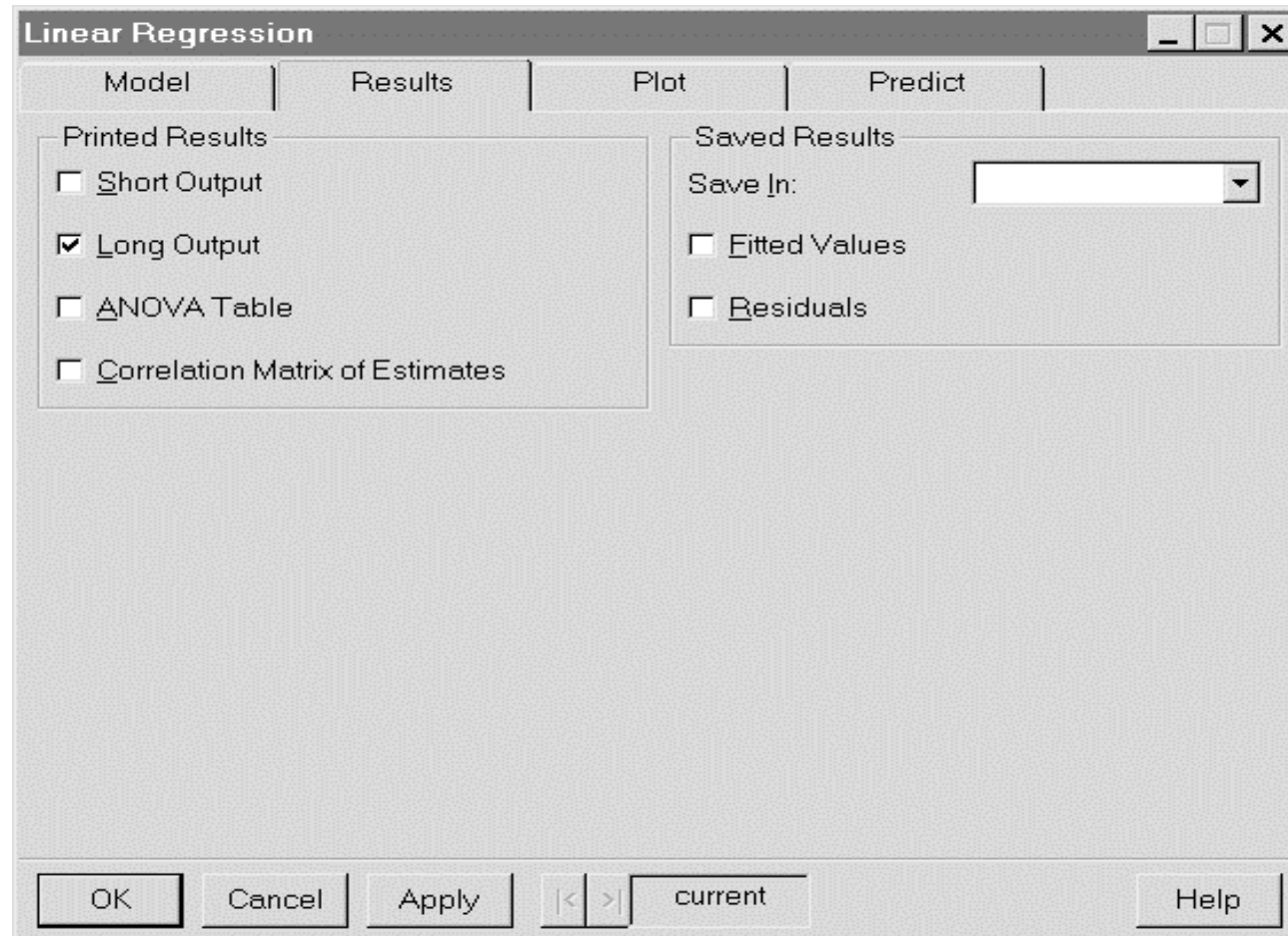
Independent: <ALL>  
age  
mass

Formula:

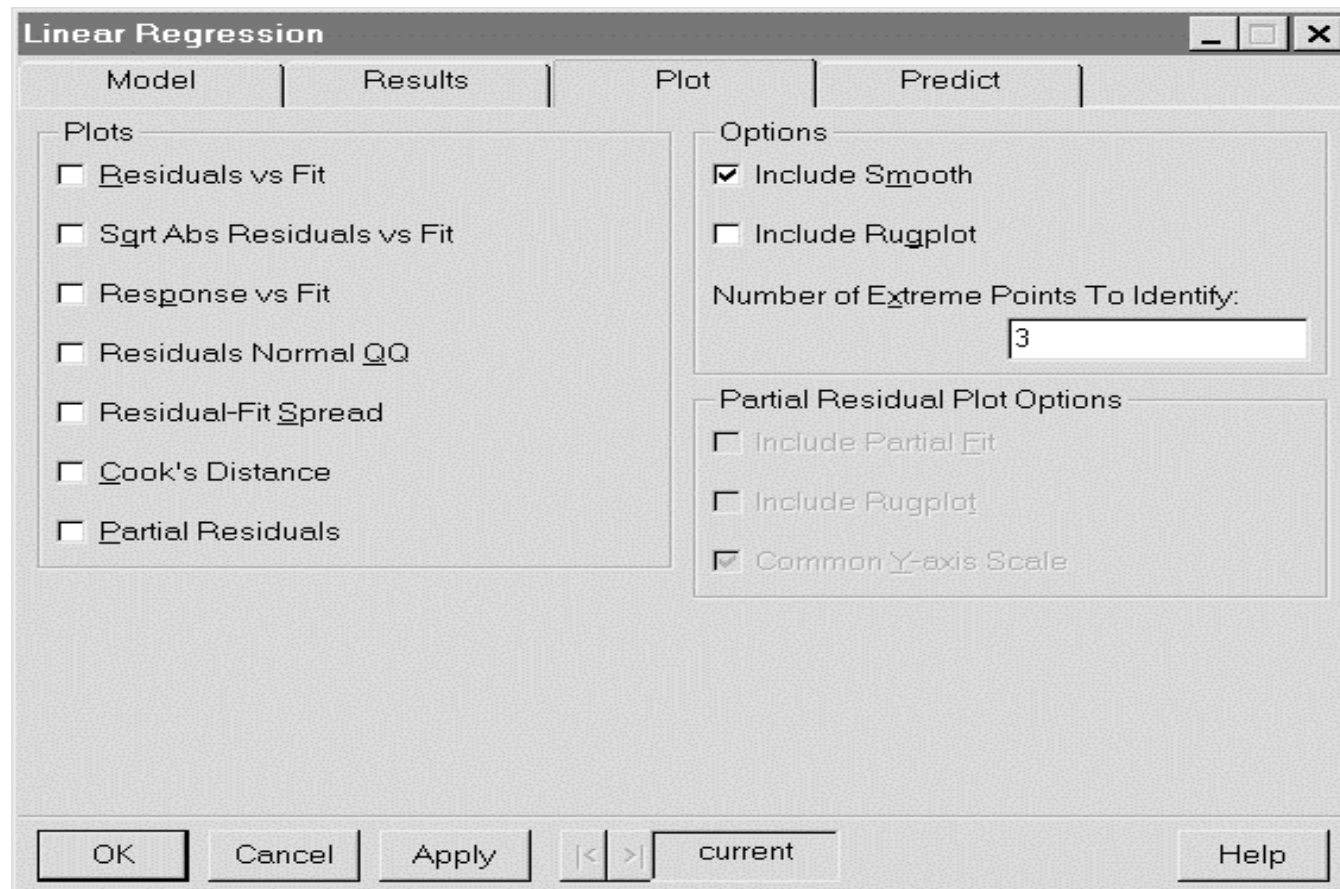
Create Formula

OK Cancel Apply |< >| current Help

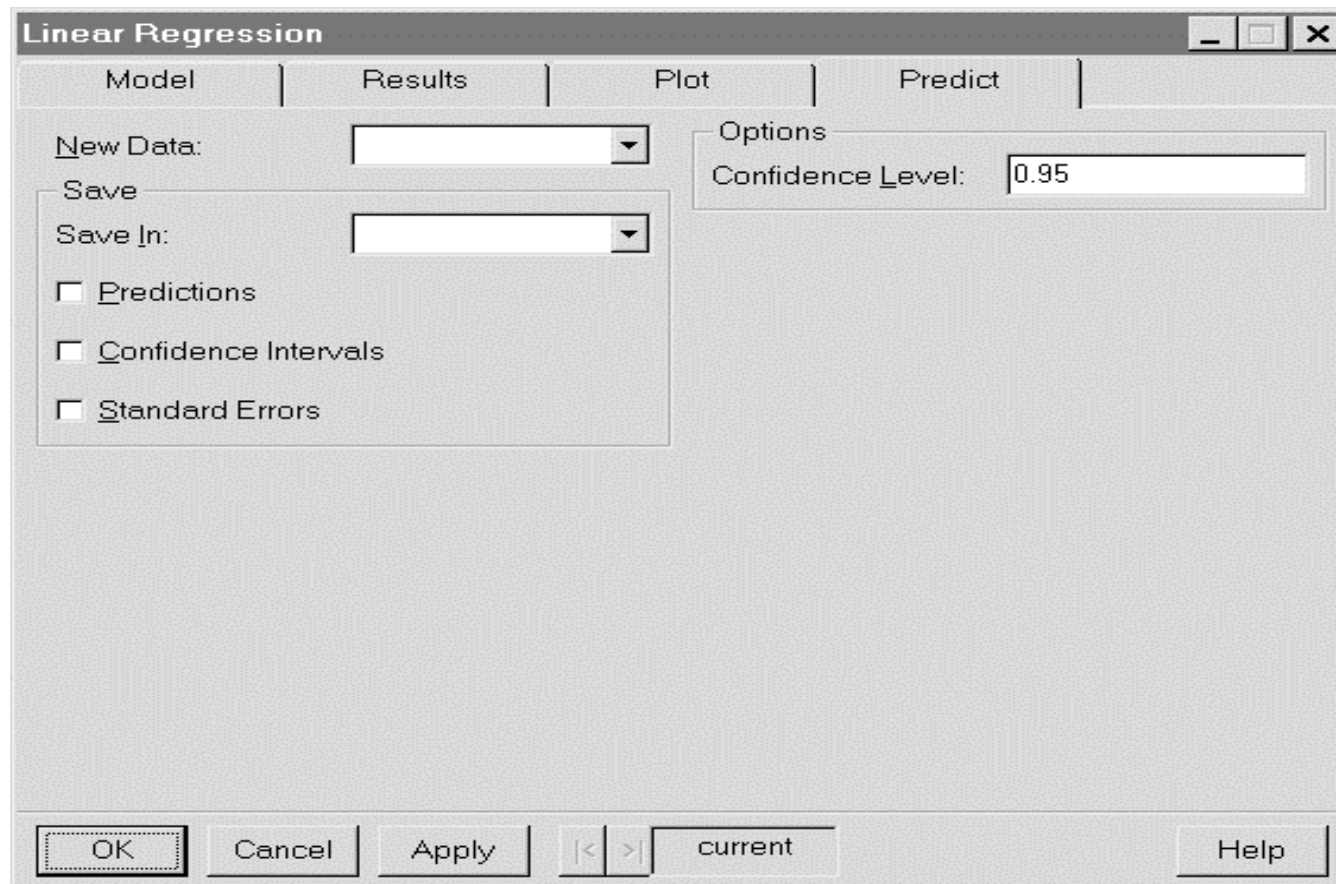
# Linear Regression Dialog : *Results*



# Linear Regression Dialog : *Plot*



# Linear Regression Dialog : *Predict*



# Linear Regression Dialog: *Model*

- Data
  - Data Frame
    - Select a data frame.
  - Weights
    - Enter the column that specifies weights to be applied to all observations used in the linear regression. To weigh all rows equally, leave this blank.

## – Subset Rows with

- Enter an S-PLUS expression which identifies the rows to use in the analysis. To use all the rows in the data frame, leave this field blank. The expression must evaluate to a vector of logical values (TRUE values are used, FALSE values are dropped), or a vector of indices identifying the numbers of the rows to use.
- Examples:
  - Species == 'bear'                      only bears are used.
  - 1:20    only the first 20 rows of the data are used.
  - Age >= 13 & Age < 20                  only teenagers are used.

- Omit Rows with Missing Values
  - Check this box to omit from the analysis any rows in the data frame that contain missing values for any of the variables in the model.
  - If this box is not checked, S-PLUS will report an error and halt the routine if any row is found to have a missing value in any of the terms in the model.
- Save Model Object
  - Save As
    - Enter the name for the object in which to save the results of the analysis. If an object with this name already exists, its contents will be overwritten.

- Formula
  - Formula
    - Enter a formula specifying the desired model. The formula specifies which regression model is to be fit. In its simplest form a formula consists of the response variable, a tilde (~), and a list of predictor variables separated by "+"s. An intercept is automatically included by default
  - Create Formula
    - Click this to open a formula builder dialog used to construct a formula specifying the desired model.

# The Formula Builder

- Formulas are required for statistical functions for
  - modeling and defining the response and explanatory variable
  - defining the structure of the model
- The formula builder can be used to build the following different models

Expression	Meaning
$T \sim F$	T is predicted linearly in F
$F_a + F_b$	Include both $F_a$ and $F_b$ in the model
$F_a - F_b$	Include all of $F_a$ except what is in $F_b$ in the model
$F_a : F_b$	The interaction between $F_a$ and $F_b$
$F_a * F_b$	$F_a + F_b + F_a : F_b$
$F_b \%in\% F_a$	$F_b$ is nested within $F_a$
$F_a / F_b$	$F_a + F_b \%in\% F_a$
$F^m$	All terms in F crossed to order m

# Formula Dialog

The image shows a 'Formula' dialog box with the following components:

- Variable** section: A label 'Choose Variable(s)' is next to a list box containing 'age' and 'mass'. Below this is a 'Transformation' button.
- Add** section: A group of six buttons for adding terms: 'Response', 'Main Effect: (+)', 'Interaction: (:)', 'Main+Interact.: (\*)', 'Quadratic: (x^2)', and 'Cubic: (x^3)'.
- Special Term** section: Three dropdown menus for 'Term Category', 'Option', and 'Format', followed by an 'Add' button.
- Remove** section: A checkbox labeled 'Remove Intercept' and a 'Term' dropdown menu, followed by a 'Remove' button.
- Formula** section: A large text input field for the formula.
- Buttons**: At the bottom are 'OK', 'Cancel', 'Apply', navigation buttons '<' and '>', a 'current' label, and a 'Help' button.

# Linear Regression Dialog: *Results*

- Printed Results
  - Short Output
    - Check here to display a short summary of the model fit. This includes
      - the model formula
      - the regression coefficients
      - the residual standard error
      - the degrees of freedom.

## – Long Output

- Check here to display a detailed summary of the model fit. This includes
  - the model formula
  - a five number summary of the residuals
  - the coefficients; their standard errors, t-statistics and p-values
  - the residual standard error and the degrees of freedom
  - the multiple R-Squared value
  - and the F test for the overall model with its degrees of freedom and p-value.

## – ANOVA Table

- Check here to display an analysis of variance table. The sums-of-squares in the table are for the terms added sequentially (Type I sums-of-squares).

## – Correlation Matrix of Estimates

- Check here to display the correlation matrix of the regression coefficients. This is only available if the Long Output is selected.

- Saved Results

- Save In

- Enter the name of an S-PLUS data frame in which fitted values and residuals of the analysis are to be saved.
      - If an object with the name you enter does not already exist, then it will be created.
      - If you enter the name of a data frame that already exists and this data frame has the same number of rows as the number of observations used in the model fit, then the saved values are appended to this data frame.
        - » This allows you to keep fitted values from a model with the original data or to keep the residuals from a number of different models for the same data in one data frame.
      - If you give the name of an existing S-PLUS object that is not a data frame or is not the appropriate size, then a warning is issued and a modified name is used.
      - You may want to specify the same data frame as on the Model page. This allows easy plotting of the fitted values or residuals with the original data.

## – Fitted Values

- Check this to save the fitted values from the model in the object specified in Save In.

## – Residuals

- Check this to save the residuals from the model in the object specified in Save In. These are the ordinary residuals; the response minus the fitted value.

# Linear Regression Dialog: *Plot*

- Plot Page
  - Plots
    - Residuals vs Fit
      - Check this to display a plot of the residuals versus the fitted values.
      - This plot is useful for checking the fit of the model, looking for outliers, and for checking the assumption of constant variance.
    - Sqrt Abs Residuals vs Fit
      - Check this to display a plot of the square root of the absolute values of the residuals versus the fitted values. This plot is useful for checking for the constant variance assumption of the model.
      - This plot is useful for looking for outliers and checking the assumption of constant variance.

- Response vs Fit
  - Check this to display a plot of the response variable versus the fitted values. The line  $y = x$  is also drawn on the graph.
  - This plot is useful for checking the fit of the model
- Residuals Normal QQ
  - Check this to display a Normal quantile-quantile plot of the residuals.
  - This plot is useful for checking the assumption of normality distributed errors
- Residual-Fit Spread
  - Check this to display a residual-fit spread plot.
  - This is a visual analog of the multiple R-squared statistic, which measures the proportion of the variability about the mean is “explained” by the fit.
  - Really it is two plots: a quantile plot of the centered values (fitted minus their mean) and a quantile plot of the residuals.

- It compares the spread of the fitted values to the spread of the residuals.
  - » If the spread of the fitted value is large relative to the spread in the residuals, this shows that the fit has “important” variable in it
  - » If the spread of the fitted value is almost as large as the spread in the residuals, this shows that the fit is not helping much to predict the response variable.

- Cook's Distance

- Check this to display a plot of Cook's distance values versus the observation number.
- This plot is useful for identifying potential outliers.
- A comparison is made between the full model and the model that lacks the observation, for each observation in the data set.
- The vertical lines on the plot represent the difference between the two models. The longer the line, the more influence that point has on the estimates.

- Plot Partial Residuals
  - Check this to display partial residual plots for all the terms in the model. A partial residual plot is a plot of  $r_i + b_k x_{ik}$  versus  $x_{ik}$  where  $r_i$  is the ordinary residual for the  $i$ -th observation,  $x_{ik}$  is the  $i$ -th observation of the  $k$ -th predictor and  $b_k$  is the regression coefficient estimate for the  $k$ -th predictor.

- Options
  - Include Smooth
    - Check this to display a smooth curve, computed with `loess.smooth`, on the Residuals vs Fit, Sqrt Abs Residuals vs Fit, and Response vs Fit plots. See the on-line help for `loess.smooth` for details.
  - Include Rugplot
    - Check this to display a rugplot on the Residuals vs Fit, Sqrt Abs Residuals vs Fit, and Response vs Fit plots. A rugplot is a sequence of vertical bars along the x-axis that mark the "observed" x values.

- Number of Extreme Points to Identify
  - Enter the number of extreme points that will be identified on the Residuals vs Fit, Sqrt Abs Residuals vs Fit, Residuals Normal QQ, and Cook's Distance plots. The row names from the data frame specified on the model page will be used to identify the points.

- Partial Residual Plots Options
  - Include Partial Fit
    - Check this to include the partial fit for the term on the plot.
  - Include Rugplot
    - Check this to display rugplots on the partial residual plots. A rugplot is a sequence of vertical bars along the x-axis that mark the "observed" x values.
  - Common Y-axis Scale
    - Check this to give all the partial residual plots the same vertical units. This is essential for comparing the importance of fitted terms in additive models.

# Linear Regression Dialog : *Predict*

- New Data
  - Enter the name of a data frame to use for computing predictions. It must contain the same names as the terms in the right side of the formula for the model. If omitted, the original data are used for computing predictions.
- Save
  - Save In
    - Enter the name of an S-PLUS data frame in which predictions, confidence intervals and standard errors are to be saved.

## – Predictions

- Check this to save the predictions in the data frame specified in Save In.

## – Confidence Intervals

- Check this to store lower and upper confidence limits in the object specified in Save In. The column names will be "N% L.C.L." and "N% U.C.L." where N is 100 times the value specified in Confidence Level. These confidence limits for the mean response are computed as the prediction plus or minus t-value times standard error.

## – Standard Errors

- Check this to store the pointwise standard errors for the predictions in the object specified in Save In.

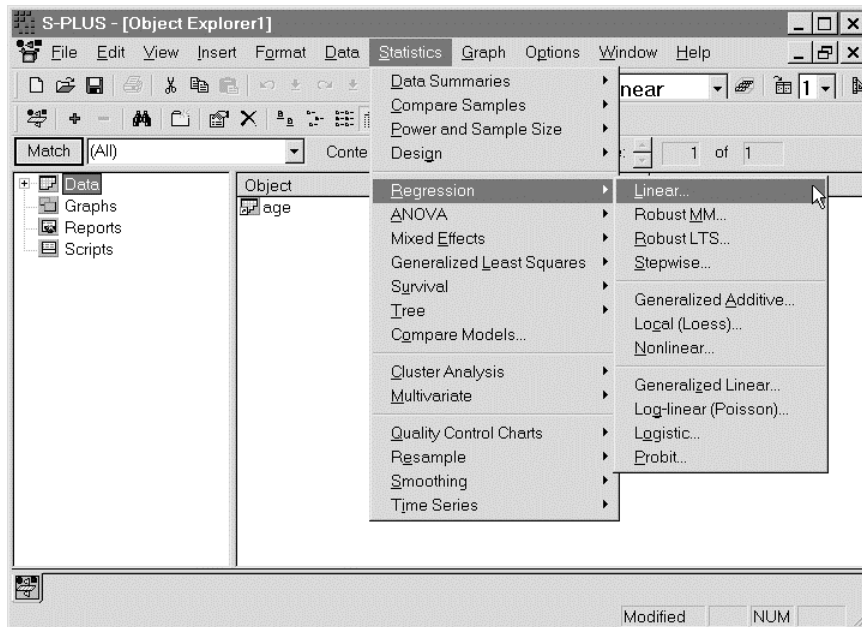
- Options
  - Confidence Level
    - Enter the confidence level to use when computing confidence intervals. This value should be less than 1 and greater than 0.
- S-PLUS language functions related to Linear Models:
  - `lm`, `plot.lm`, `predict.lm`, `print.lm`, `summary.lm`, `aov`, `gam`, `glm`, `loess`, `nls`

# Example:

Fit a linear model for the mass data.

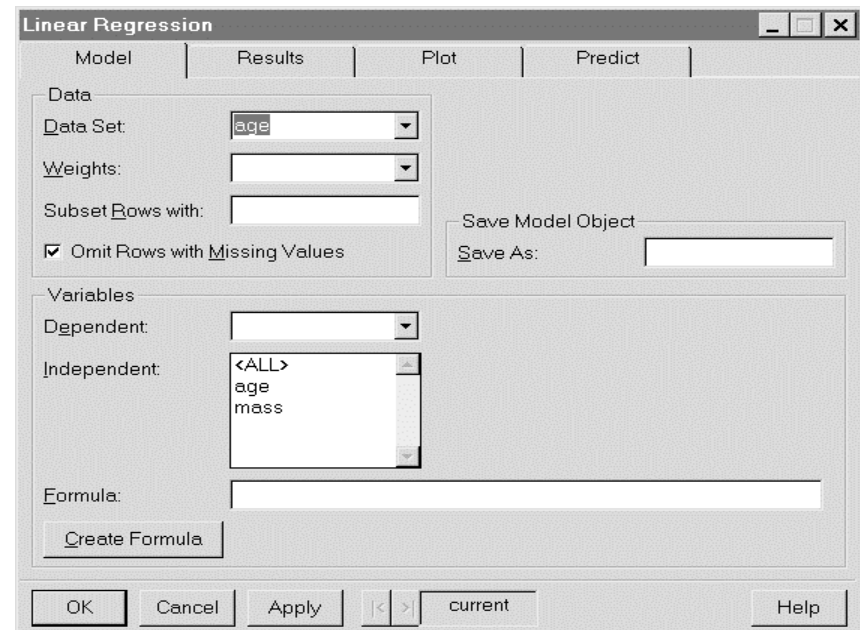
## Step 1

Open the linear  
Regression Dialog



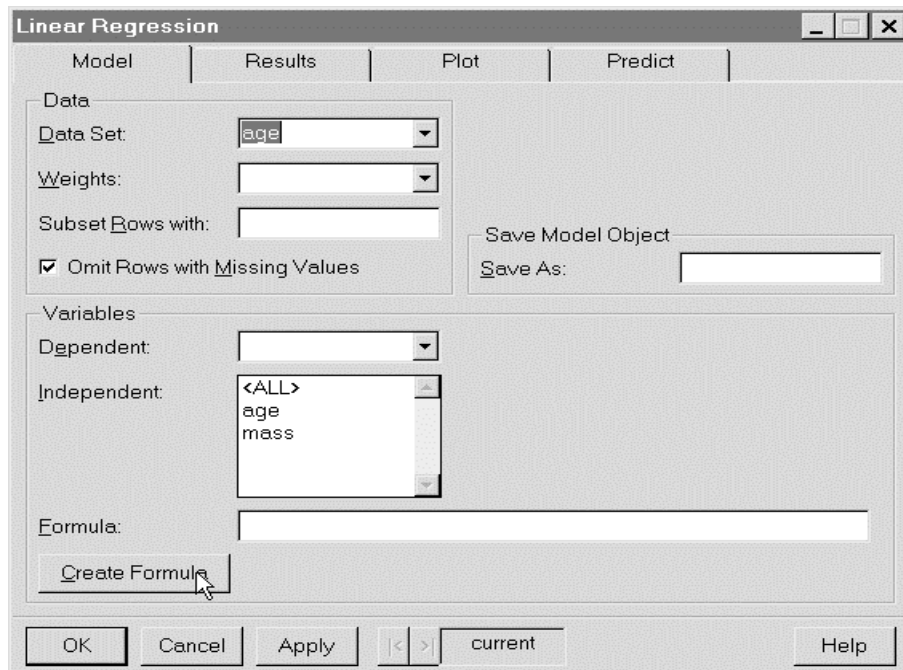
## Step 2

Fill the text field for  
the Data Frame



**Step 3**  
Click on Create  
Formula

**Step 4 \***  
Select the response  
and the predictor variables



The Linear Regression dialog box has four tabs: Model, Results, Plot, and Predict. The Model tab is active. It contains sections for Data, Variables, and Formula. In the Data section, 'age' is selected for the Data Set. In the Variables section, 'age' and 'mass' are listed under Independent. The Formula section has a 'Create Formula' button. At the bottom are OK, Cancel, Apply, and Help buttons.

Linear Regression

Model Results Plot Predict

Data

Data Set: age

Weights:

Subset Rows with:

☒ Omit Rows with Missing Values

Save Model Object

Save As:

Variables

Dependent:

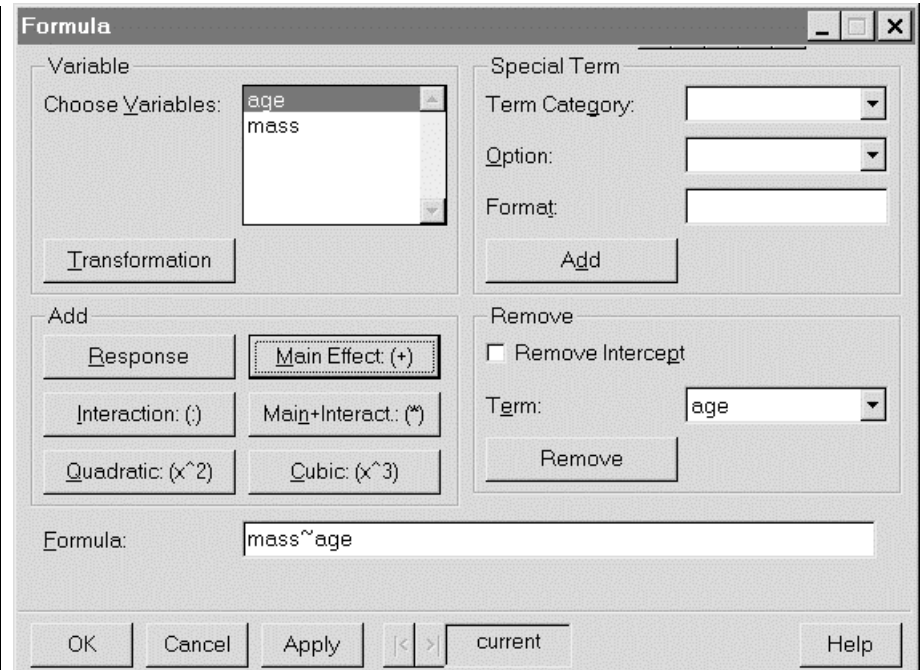
Independent:

<ALL>  
age  
mass

Formula:

Create Formula

OK Cancel Apply |< >| current Help



The Formula dialog box is used to build the regression formula. It includes a list of variables (age, mass), buttons for adding terms (Response, Main Effect, Interaction, Quadratic, Cubic), and options for special terms and intercept. The formula 'mass~age' is entered in the Formula field. At the bottom are OK, Cancel, Apply, and Help buttons.

Formula

Variable

Choose Variables: age  
mass

Transformation

Add

Response Main Effect: (+)

Interaction: () Main+Interact.: (\*)

Quadratic: (x^2) Cubic: (x^3)

Special Term

Term Category:

Option:

Format:

Add

Remove

☐ Remove Intercept

Term: age

Remove

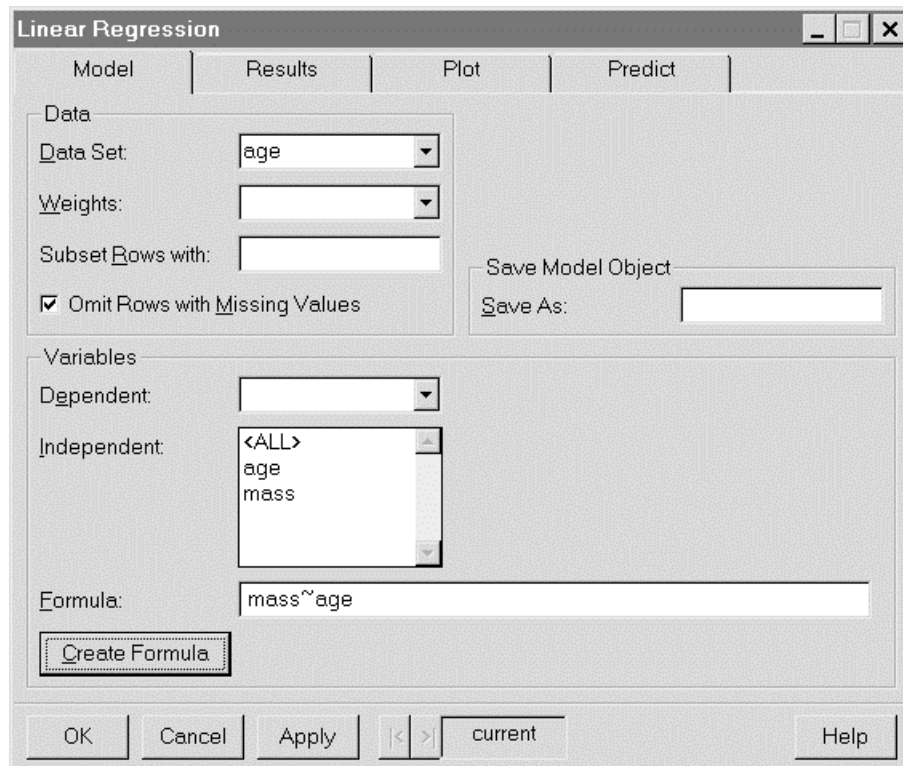
Formula: mass~age

OK Cancel Apply |< >| current Help

\*

- To select the response variable of the model, select **mass** from the **choose Variables** list box
- Click on the button **Response** to set the chosen variable as the response
- Select the predictor variable, which is **age** from the **choose Variables** list box
- Set the predictor by clicking on the button **Main Effect:(+)**
- Click OK in the formula Builder.

Step 5  
The Model Dialog  
should look like this



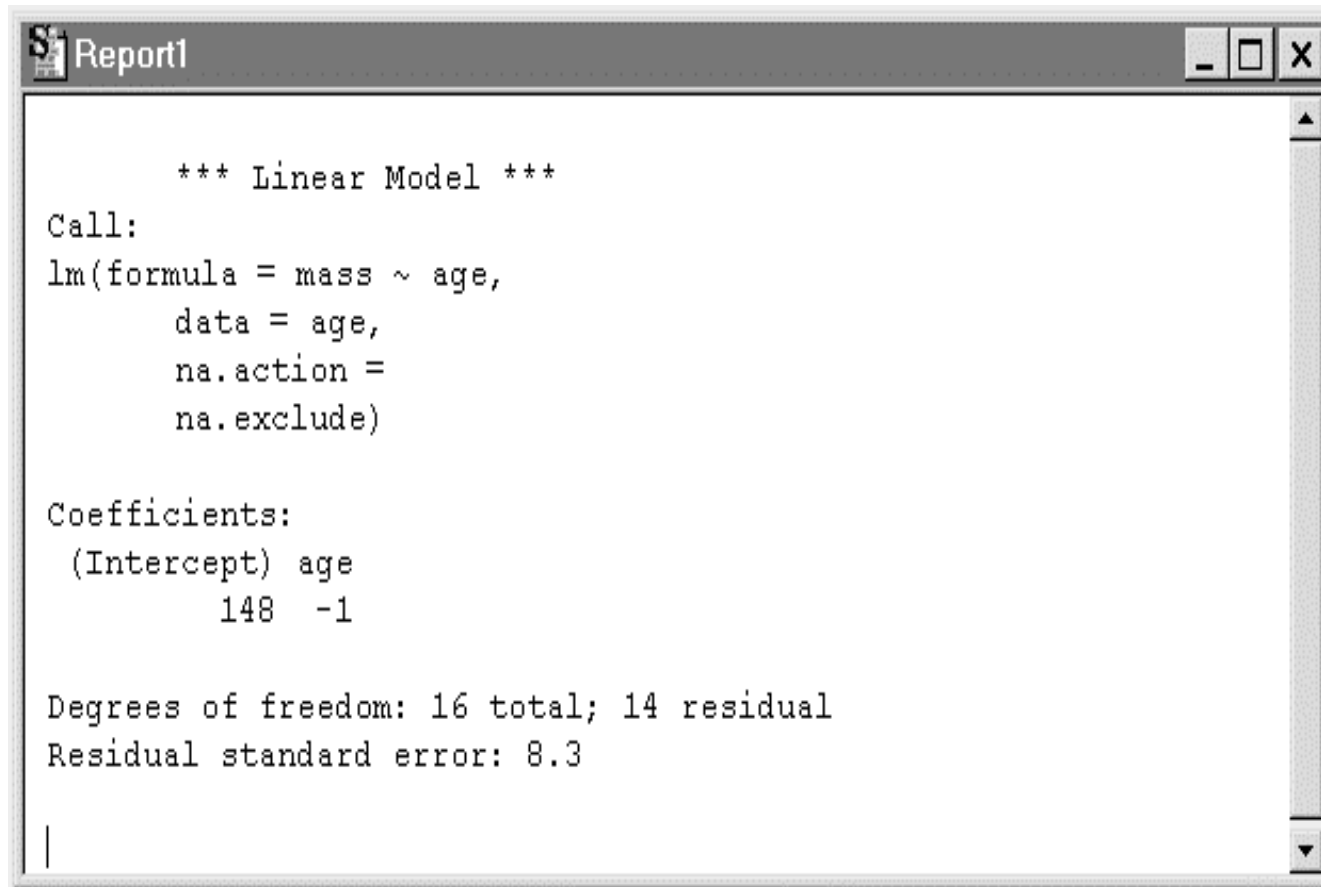
The screenshot shows the 'Linear Regression' dialog box with the following settings:

- Data:**
  - Data Set: age
  - Weights: (empty)
  - Subset Rows with: (empty)
  - ☒ Omit Rows with Missing Values
- Save Model Object:**
  - Save As: (empty)
- Variables:**
  - Dependent: (empty)
  - Independent: <ALL>, age, mass
- Formula:**
  - mass~age
  - Create Formula button

Buttons at the bottom: OK, Cancel, Apply, < >, current, Help.

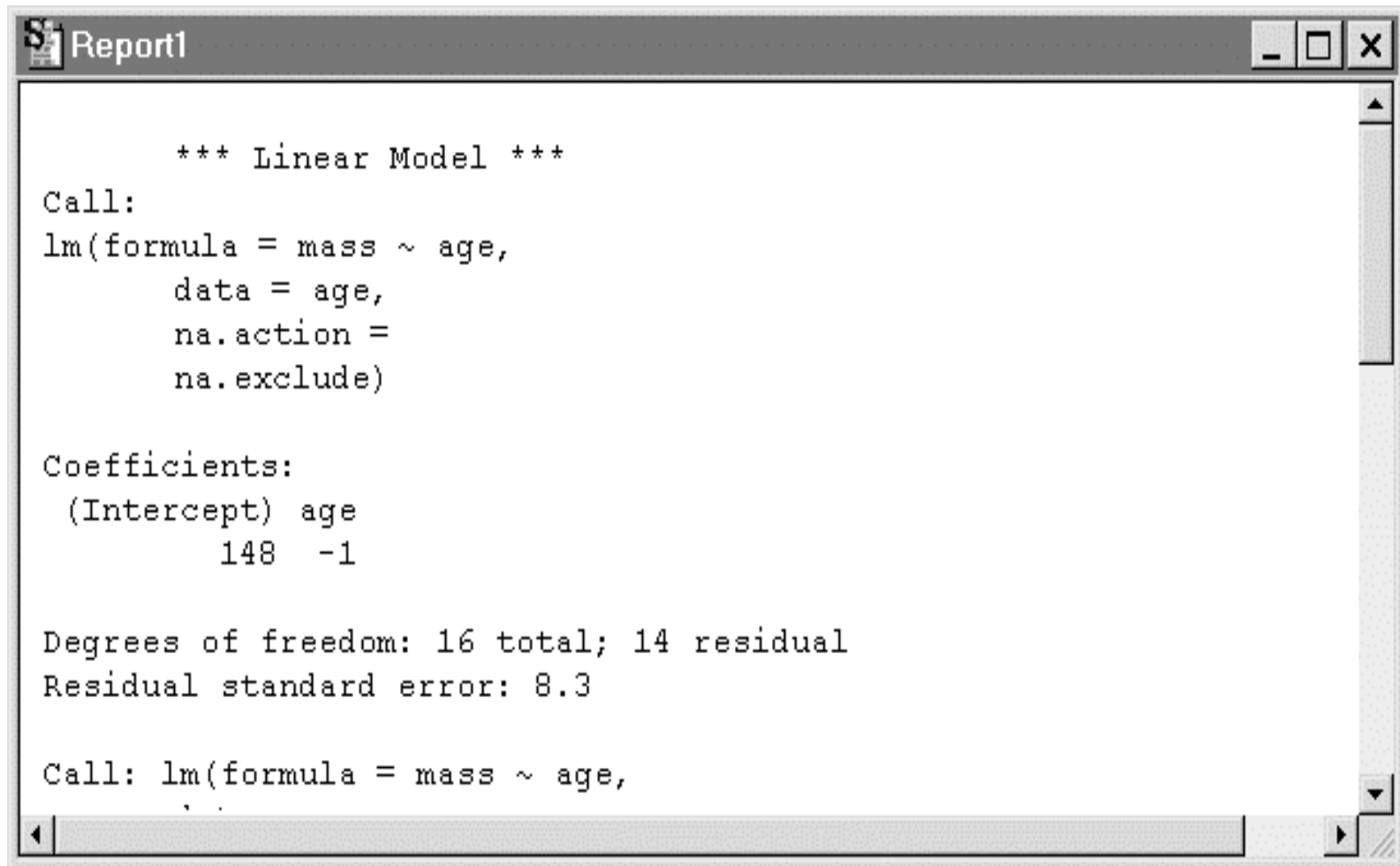
Hit Ok. The result  
depends on the  
choices you make  
in the Result,  
Plot and Predict  
Dialog

# Example: *Short Output*



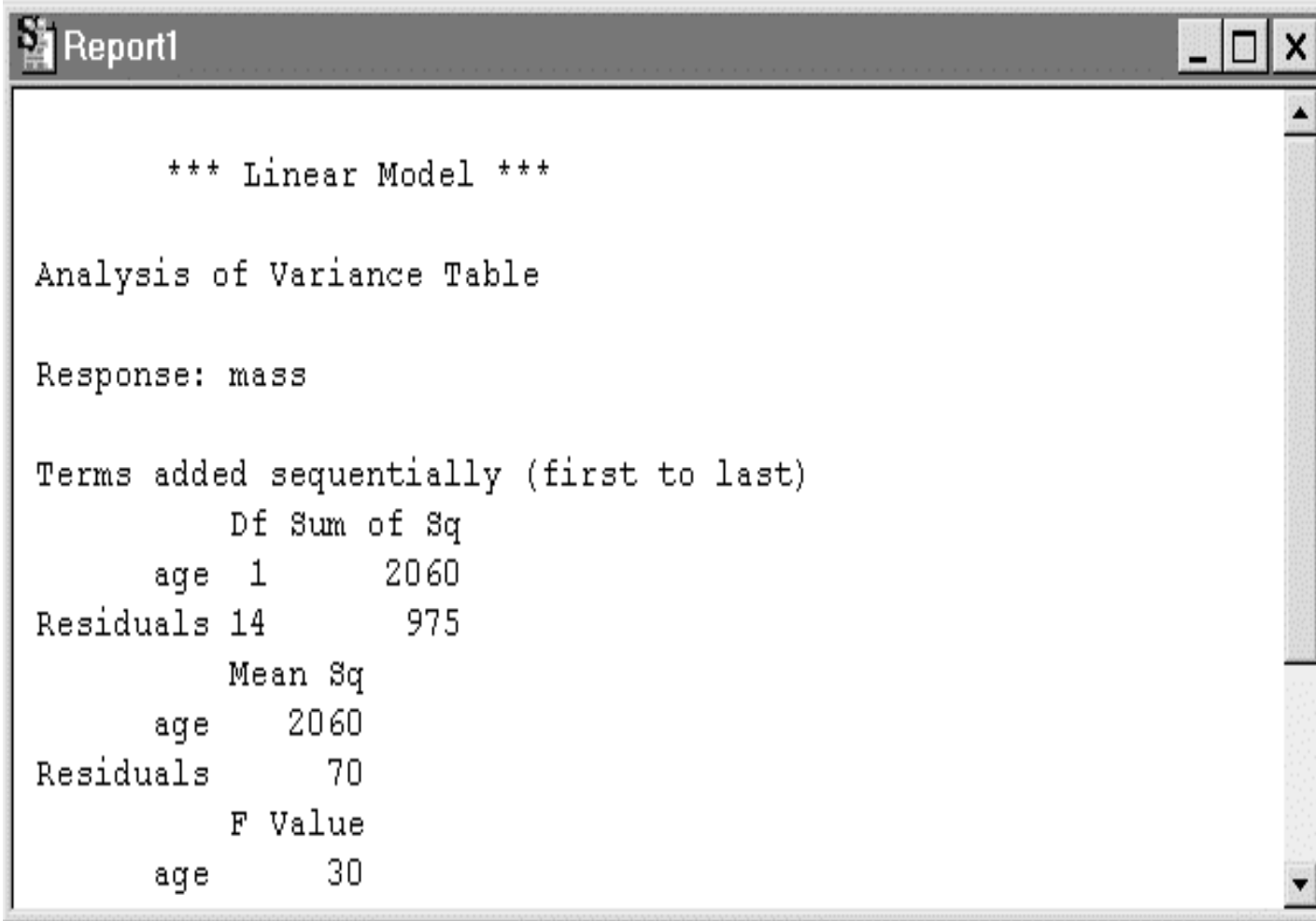
```
*** Linear Model ***  
Call:  
lm(formula = mass ~ age,  
    data = age,  
    na.action =  
    na.exclude)  
  
Coefficients:  
 (Intercept) age  
          148  -1  
  
Degrees of freedom: 16 total; 14 residual  
Residual standard error: 8.3  
|
```

# Example: *Long Output*



```
*** Linear Model ***  
Call:  
lm(formula = mass ~ age,  
    data = age,  
    na.action =  
    na.exclude)  
  
Coefficients:  
 (Intercept) age  
          148  -1  
  
Degrees of freedom: 16 total; 14 residual  
Residual standard error: 8.3  
  
Call: lm(formula = mass ~ age,
```

# Example: *ANOVA Table*



The screenshot shows a window titled 'Report1' with standard window controls (minimize, maximize, close) in the top right corner. The main content area displays the following text:

```
*** Linear Model ***  
  
Analysis of Variance Table  
  
Response: mass  
  
Terms added sequentially (first to last)  
      Df Sum of Sq  
age    1      2060  
Residuals 14      975  
  
      Mean Sq  
age      2060  
Residuals    70  
  
      F Value  
age         30
```

	Df	Sum of Sq
age	1	2060
Residuals	14	975

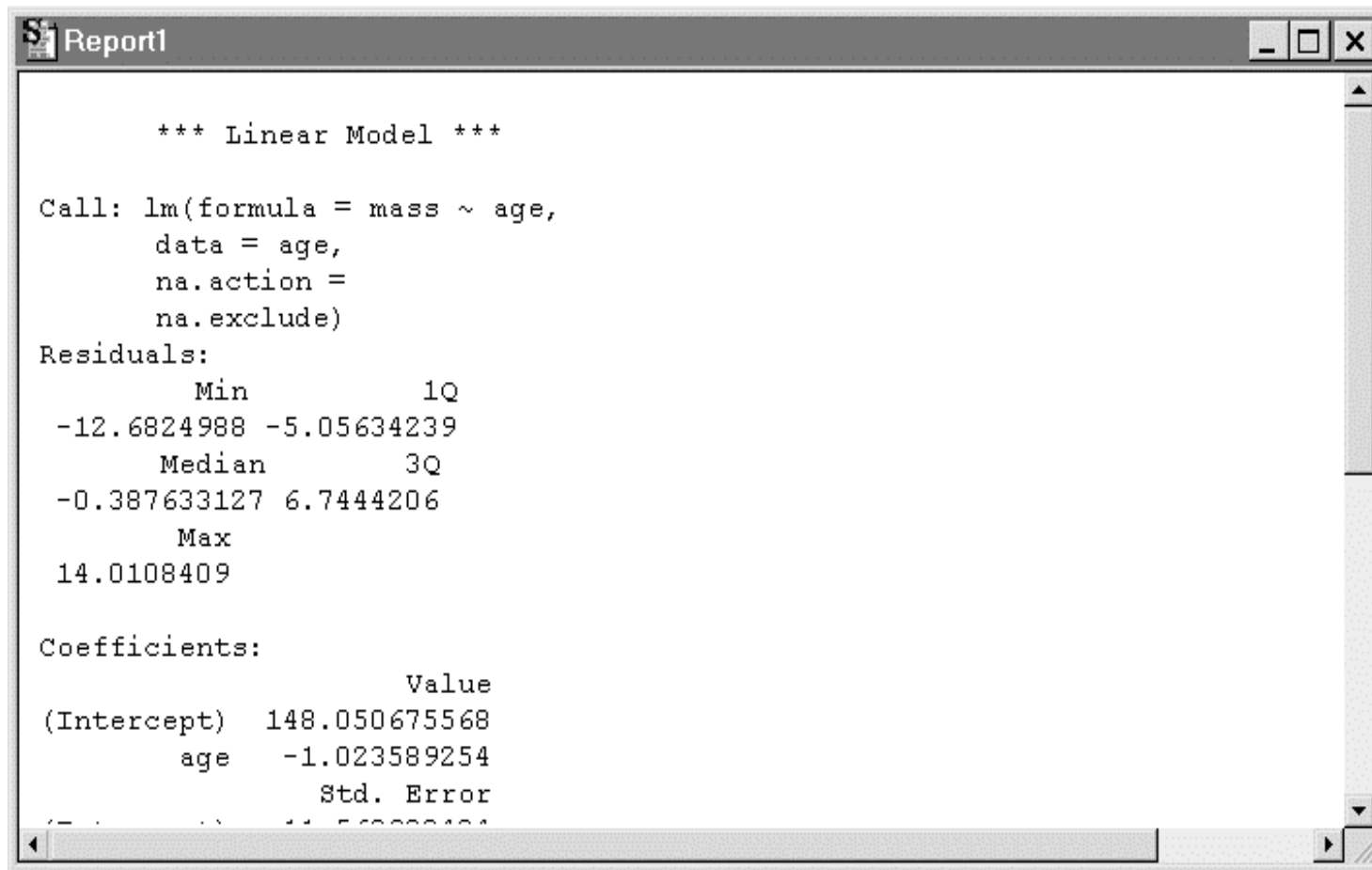
  

	Mean Sq
age	2060
Residuals	70

	F Value
age	30

# Example: *Correlation Matrix of Estimates*

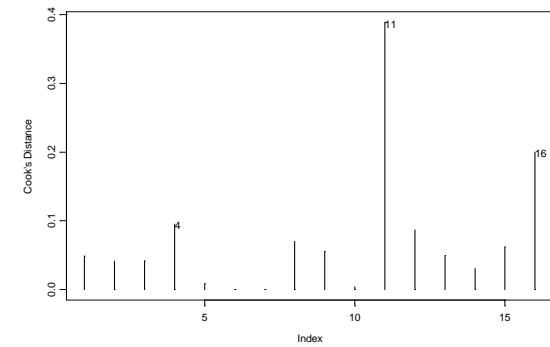
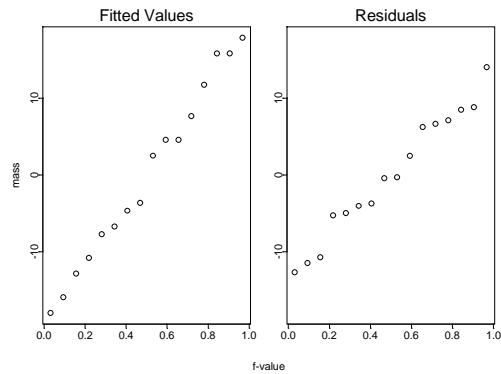
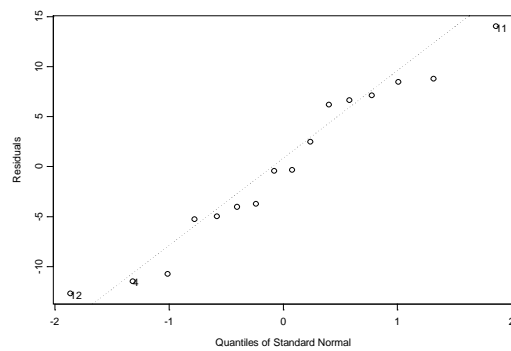
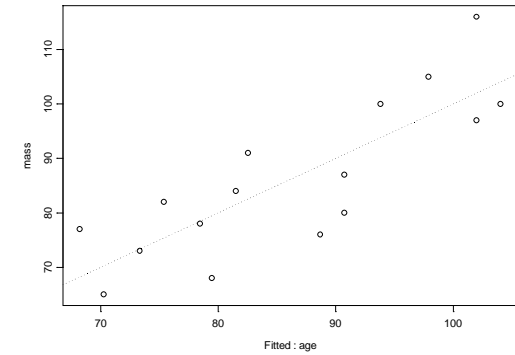
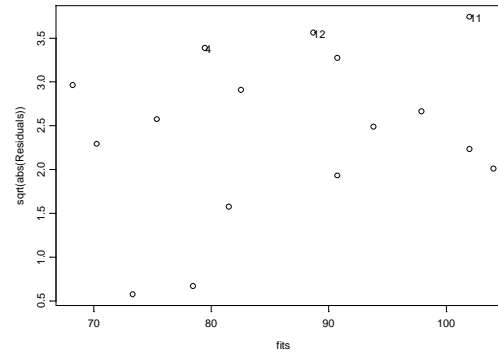
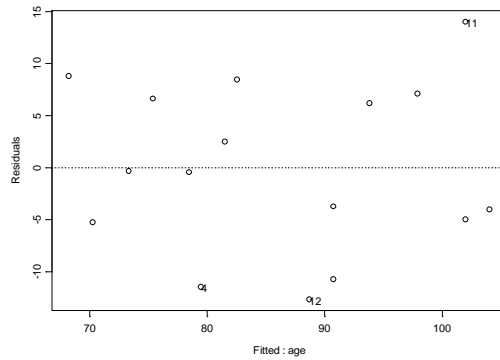


```
*** Linear Model ***

Call: lm(formula = mass ~ age,
  data = age,
  na.action =
  na.exclude)
Residuals:
      Min       1Q   -12.6824988 -5.05634239
      Median       3Q   -0.387633127  6.7444206
      Max
    14.0108409

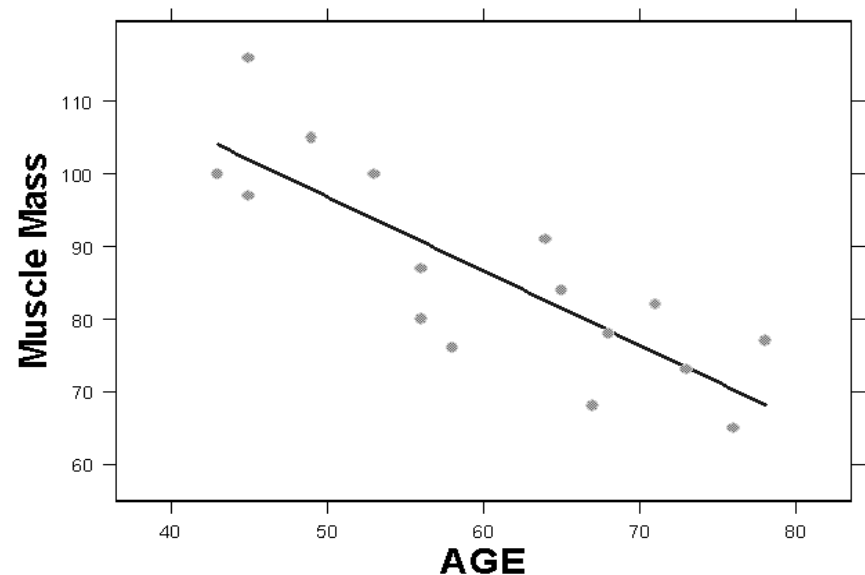
Coefficients:
              Value
(Intercept) 148.050675568
          age  -1.023589254
              Std. Error
(Intercept) 11.568888101
          age  0.145141401
```

# Example: *Plots*



# Example: *Plot of the line on the scatter plot*

- To Create the plot
  - Select the data columns
  - Open the 2D Palette
  - Click the 2D Linear fit plot



# **Exporting Data**

# Exporting Data

- Exporting data to ASCII files
- Exporting to Encapsulated Postscript Files
- Printing
- Embedding an S-Plus Graphics in Another Application
- Creating a PowerPoint Presentation

# Exporting data to ASCII files

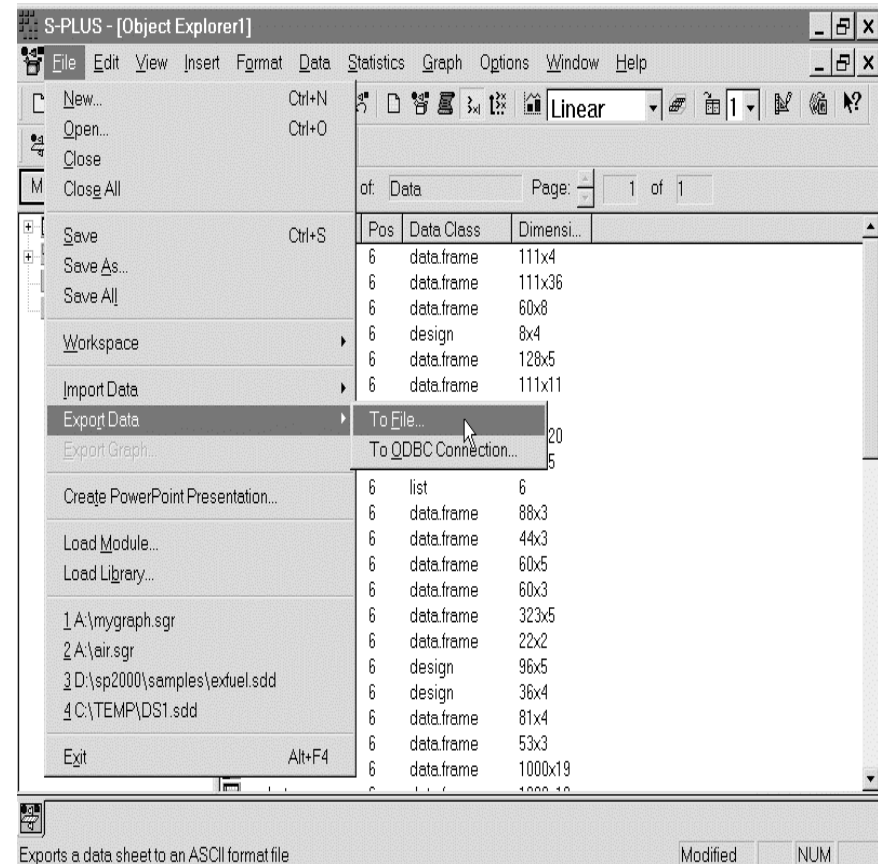
- From the “**Main Menu**” Select “**File**”
  - Select “**Export Data**”
  - Select “**To File**”
  - “**Export Data**” Dialog appears
    - In the “**Data Specs**” dialog specify the following
      - Specify a file name
      - Select a file Type here use **Text Files (\*.txt,\*.prn)**
      - Specify the data object being exported
    - In the “**Option Page**”
      - Specify the delimiter

- » by default the delimiter is the comma
- » user-defined delimiters must contain 8 or fewer characters
- » an empty field is exported with fixed length line
- Specify whether to include column names in the file
- Specify whether to have quotation marks used to enclose character data in the file.
- Click **“SAVE”**

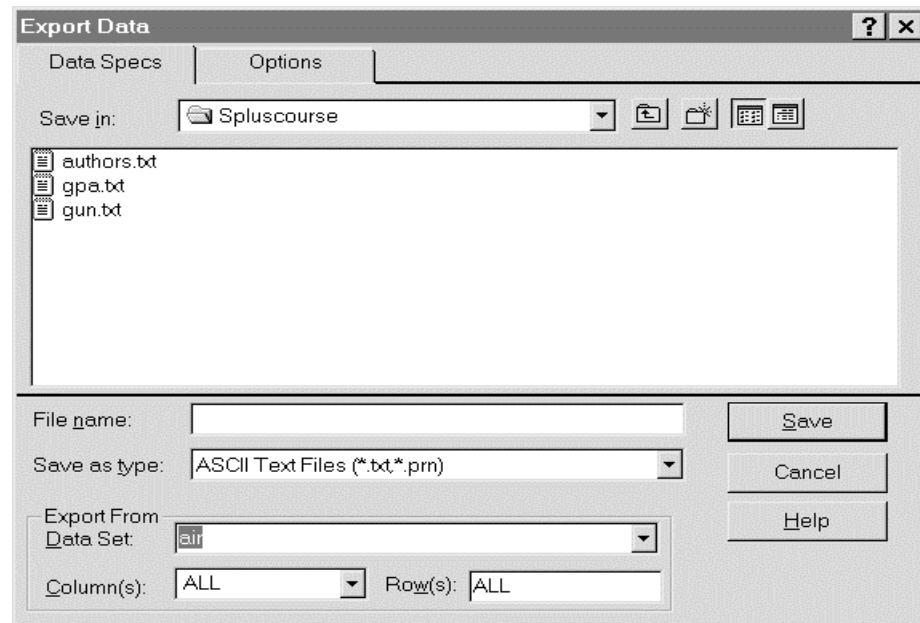
# Example: Exporting data to ASCII files

- Export air built-in-data frame to an ASCII file

**Step 1**

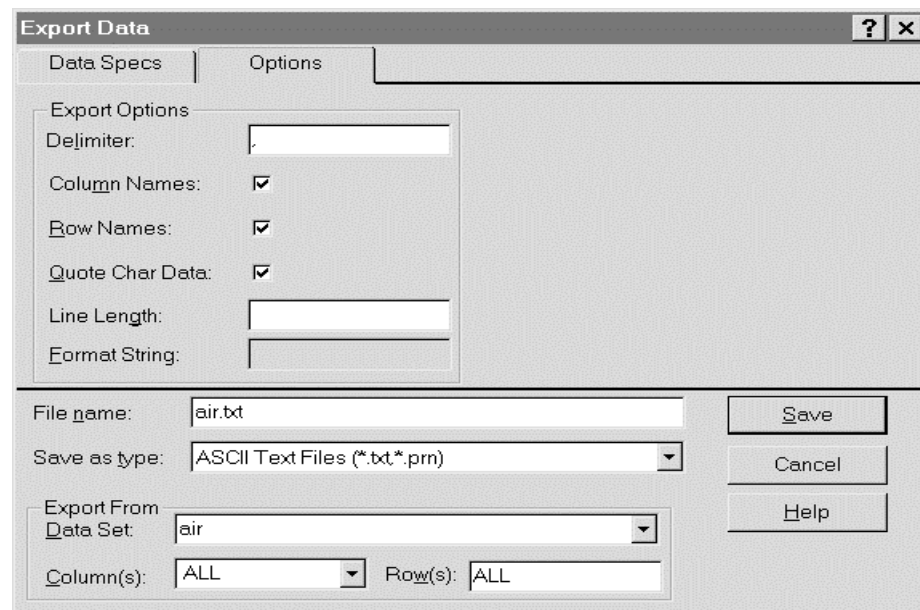


**Step 2**



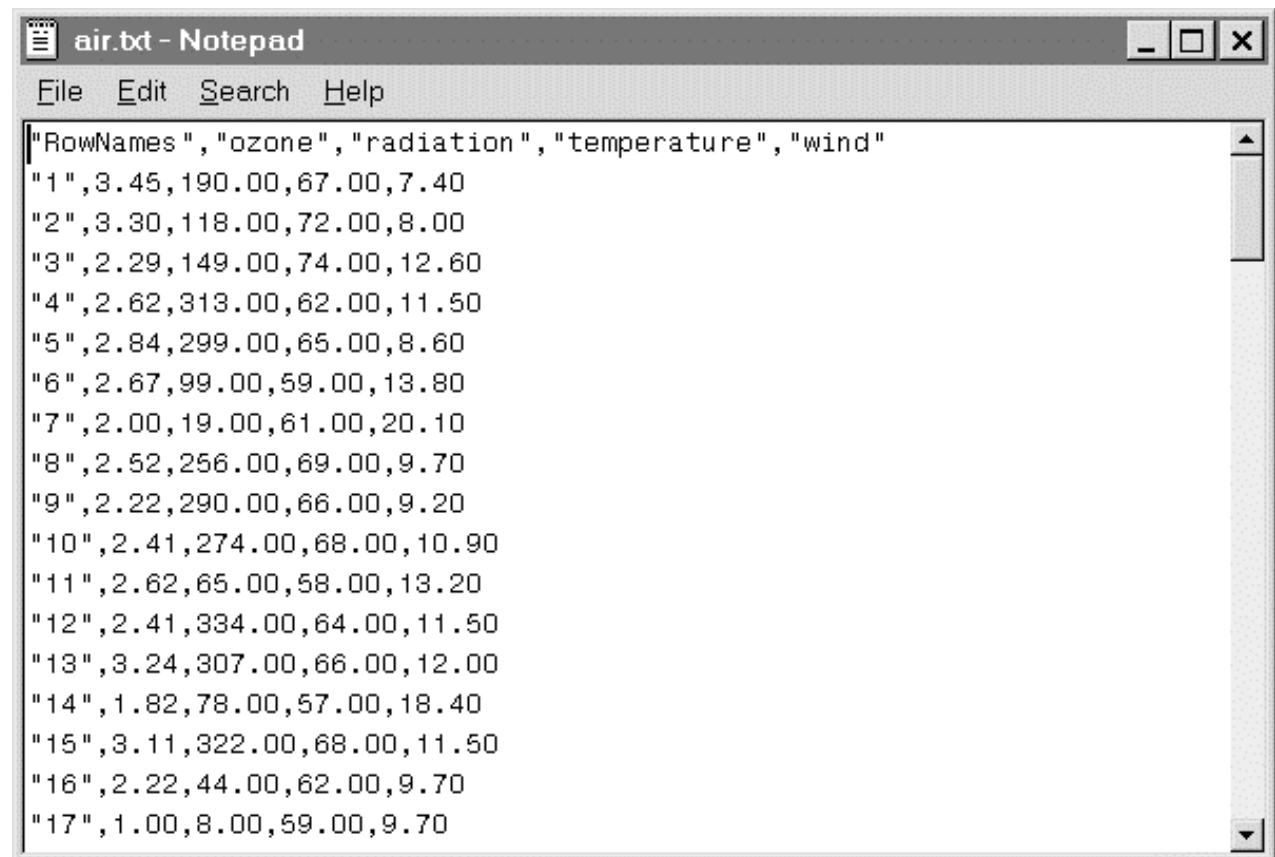
The "Export Data" dialog box is shown with the "Data Specs" tab selected. The "Save in:" field is set to "Spluscourse". A list of files is displayed: authors.txt, gpa.txt, and gun.txt. The "File name:" field is empty. The "Save as type:" is set to "ASCII Text Files (\*.txt\*.prn)". The "Export From Data Set:" is set to "air". The "Column(s):" is set to "ALL" and the "Row(s):" is set to "ALL". Buttons for "Save", "Cancel", and "Help" are on the right.

**Step 3**



The "Export Data" dialog box is shown with the "Options" tab selected. The "Export Options" section is expanded, showing "Delimiter:" (empty), "Column Names:" (checked), "Row Names:" (checked), "Quote Char Data:" (checked), "Line Length:" (empty), and "Format String:" (empty). The "File name:" field is now set to "air.txt". The "Save as type:" remains "ASCII Text Files (\*.txt\*.prn)". The "Export From Data Set:" is still "air", and "Column(s):" and "Row(s):" are still "ALL". Buttons for "Save", "Cancel", and "Help" are on the right.

**Result**

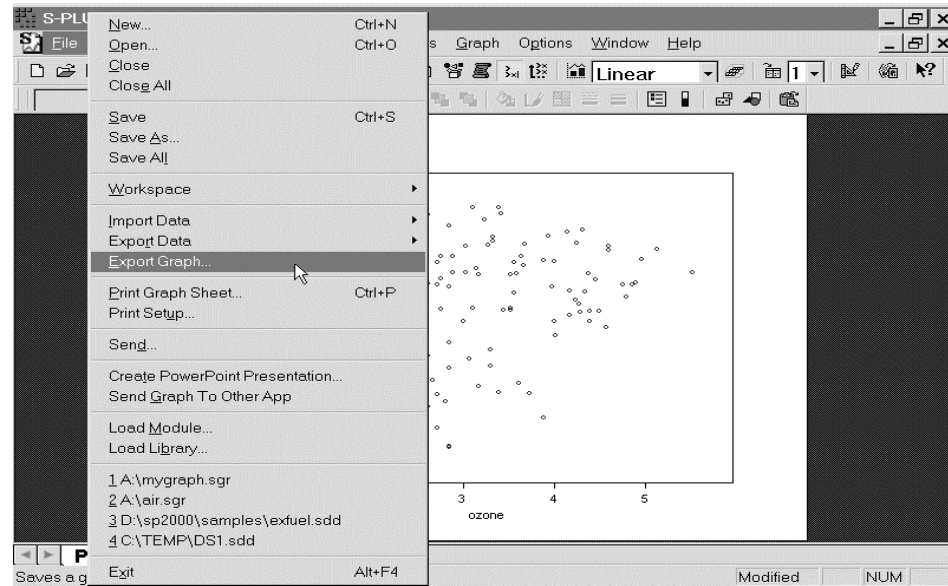


```
air.txt - Notepad
File Edit Search Help
"RowNames","ozone","radiation","temperature","wind"
"1",3.45,190.00,67.00,7.40
"2",3.30,118.00,72.00,8.00
"3",2.29,149.00,74.00,12.60
"4",2.62,313.00,62.00,11.50
"5",2.84,299.00,65.00,8.60
"6",2.67,99.00,59.00,13.80
"7",2.00,19.00,61.00,20.10
"8",2.52,256.00,69.00,9.70
"9",2.22,290.00,66.00,9.20
"10",2.41,274.00,68.00,10.90
"11",2.62,65.00,58.00,13.20
"12",2.41,334.00,64.00,11.50
"13",3.24,307.00,66.00,12.00
"14",1.82,78.00,57.00,18.40
"15",3.11,322.00,68.00,11.50
"16",2.22,44.00,62.00,9.70
"17",1.00,8.00,59.00,9.70
```

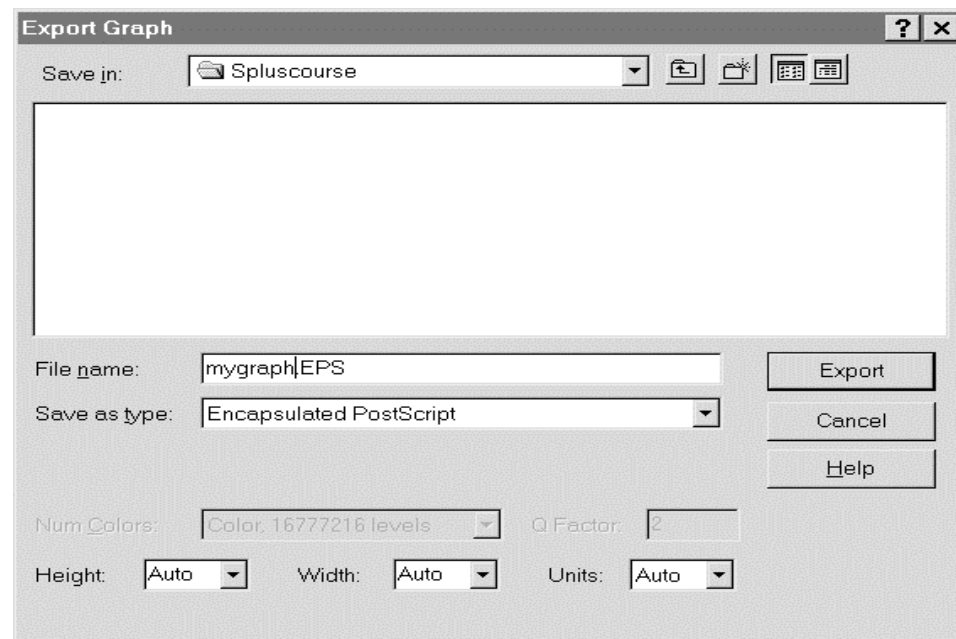
# Exporting to Encapsulated Postscript Files

- To export Graph sheet to EPS file
- From the “**Main Menu**” Select “**File**”
  - Select “**Export Graph**”
  - Select the desired file type in “**Save as Type**”
    - Here choose EPS[\* .EPS]
  - Type a file name in “**File Name**”
  - Select “**Save**” to save the sheet


**Step 1**



**Step 2**



# Printing

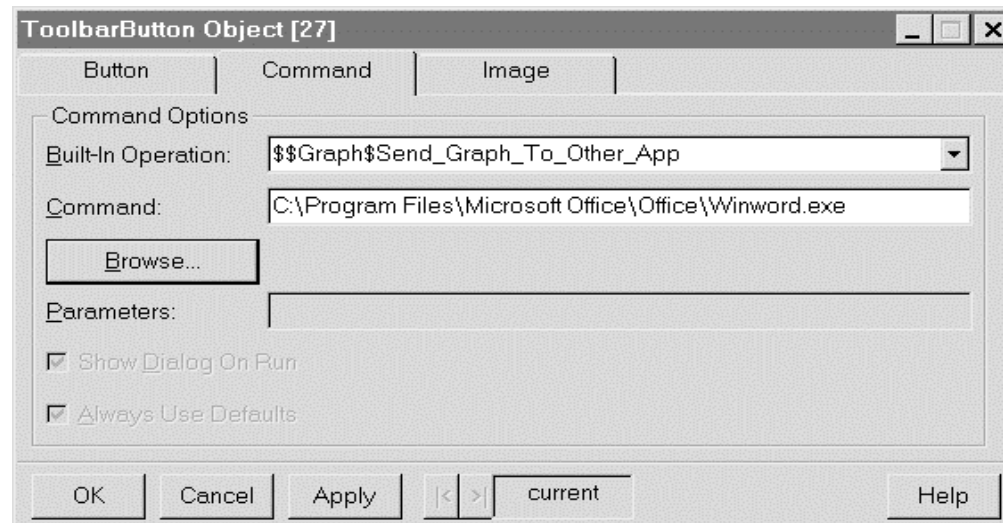
- You can print
  - S-Plus Graph Sheets
  - Data Windows
  - Report files
  - Scripts
- Using the **Print** button you can print the current selected window 
- You can also Print using the “**Print**” Dialog
  - Select “**File**”
  - Select “**Print**”
  - Select “**Ok**”

# Embedding an S-Plus Graphics in Another Application


- The “**Send Graph**” button copies the current Graph sheet to the clipboard, then switches automatically into a specific application for pasting.
- Customizing the “**Send Graph**” button
  - Right-click on the “**Send Graph**” from the “**Toolbar Button Object**” dialog

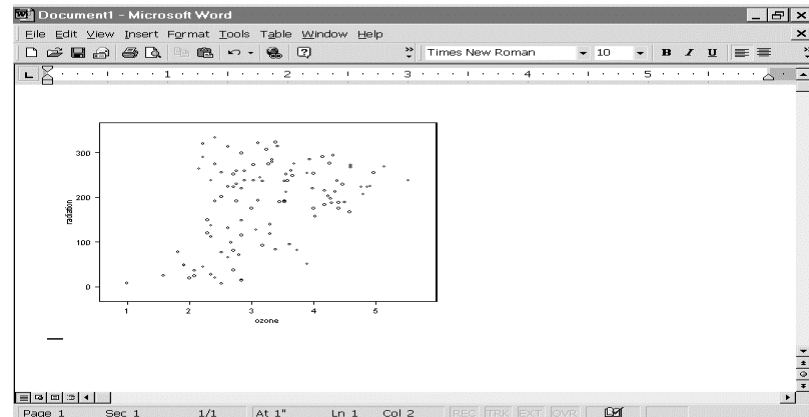


- Scroll through the Built-In Operation choose
  - \$\$Graph\$Send\_Graph\_To\_Other\_App
- Use the “**Browse**” button and Specify the complete path to the target application’s .EXE file.
  - For example, browse through the Program files of the computer to find the “**Winword.exe**” file.
- Click **OK**.
- **Note that:** the **Send Graph** button icon changes to reflect the target application



# Example: Using the “Send Graph” button

- Select the Graph sheet.
- Click on the “**Send Graph**” Button 
- The target application is, Microsoft word, is launched, and a message appears indicating the Graph sheet is ready for pasting.
- Click “OK” to confirm the message
- Paste the graph sheet in the Microsoft word document by choosing EDIT/PASTE.

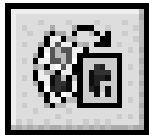


- To edit the embedded S-Plus object:
  - Double-click on the embedded S-Plus object
  - Edit the S-Plus graph by double clicking or using the right-click menu item
  - When finished editing, click anywhere outside the embedded object

# Creating a PowerPoint Presentation

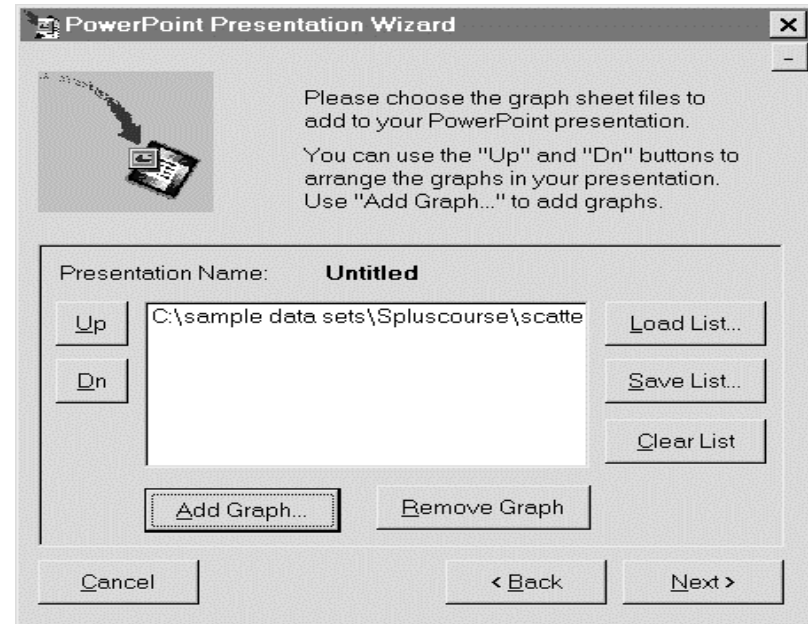
- In order to create a PowerPoint Presentation, make sure of the following
  - The PowerPoint options enabled during installation of S-Plus 4.5
  - The S-Plus graph sheets you're planning to use are saved
  - You have power point 7.0 or higher

## Example: To create a PowerPoint presentation

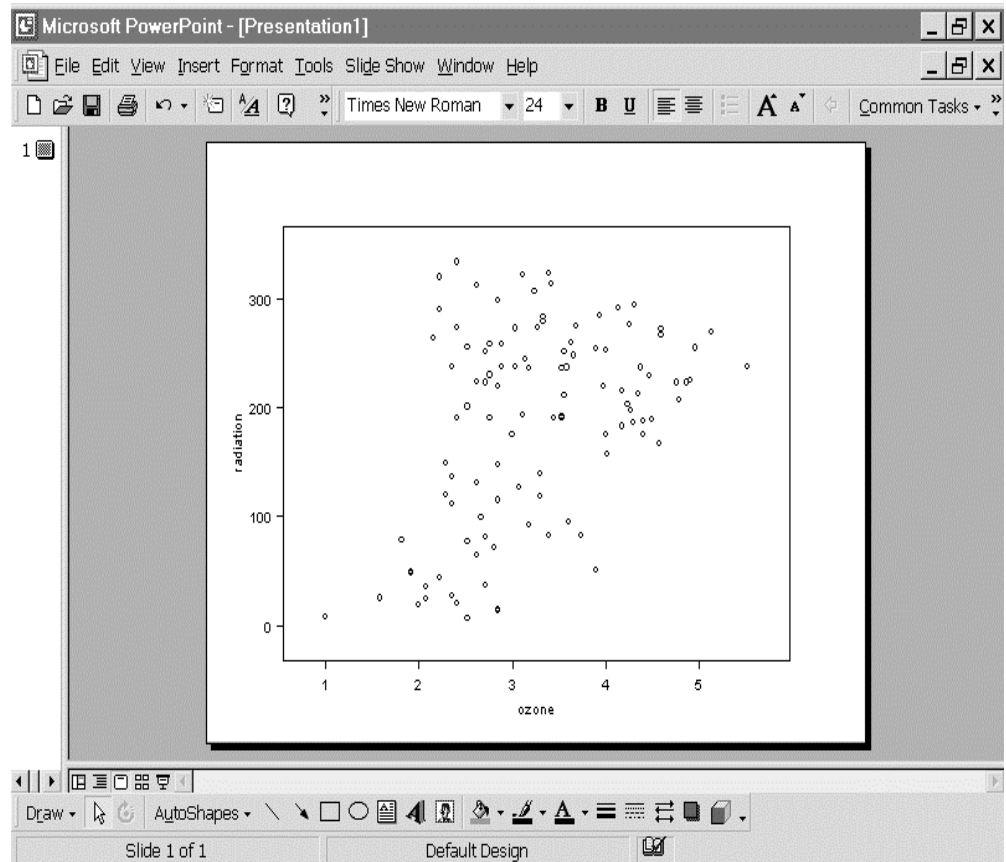
- Export the S-Plus Graph file “scatter.sgr” into a PowerPoint presentation
- Click on the PowerPoint presentation button on the standard toolbar 
- The PowerPoint presentation Wizard will startup, click “Next” on the “Welcome Screen”



- Click on the “**Add Graph**” button to find S-Plus graphs on your system that you would like to add to the list for this presentation
- Click “**Next**” to move to the next page of the wizard
- Click “**Finish**”. PowerPoint is started and the graphs you chose are inserted in slides in a new PowerPoint presentation.



- Save the file in your PowerPoint presentation before you exit the PowerPoint Wizard.



# History Log & Script Files

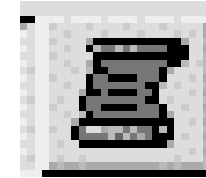
# Objectives:

- Creating Script Files from the History Log
- Linking the History Log to a Toolbar Button
- Customizing Toolbars, Menus, and dialogs
- Restoring the Default Toolbars
- Creating Custom menu Item and Dialogs

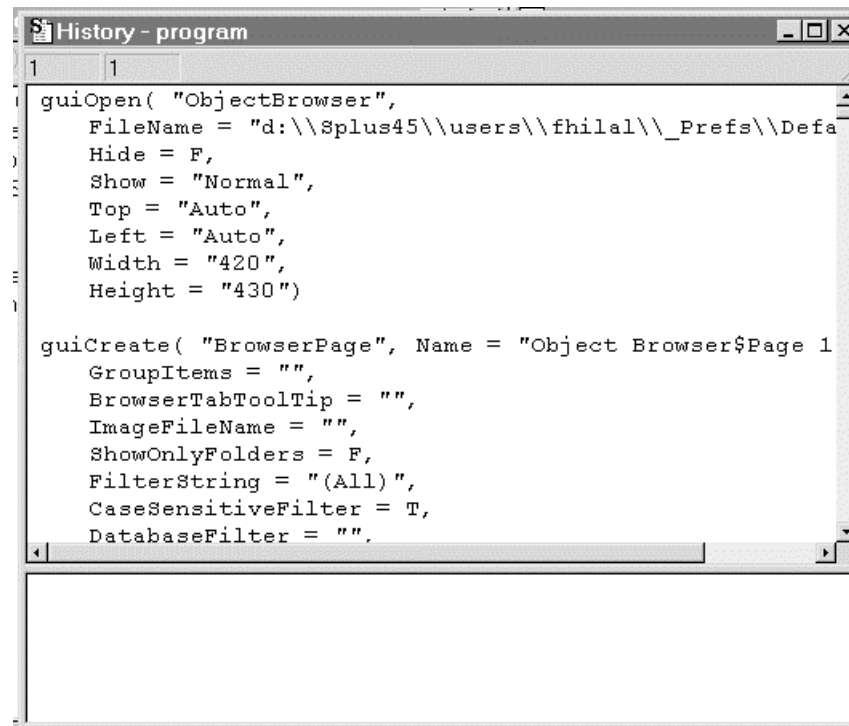
- Everything done in S-Plus is recorded behind the scenes in the S programming language.
- This code can be viewed at any time by displaying the **History Log**.
- The **History Log** can be saved edited and re-executed to automatic repetitive tasks.

# Viewing the History Log

- Click on the **History Log** button



Result

A screenshot of a MATLAB script window titled "History - program". The window shows a script with two main functions: `guiOpen` and `guiCreate`. The `guiOpen` function opens an "ObjectBrowser" window with specific file names and dimensions. The `guiCreate` function creates a "BrowserPage" with various properties like `GroupItems`, `BrowserTabToolTip`, `ImageFileName`, `ShowOnlyFolders`, `FilterString`, `CaseSensitiveFilter`, and `DatabaseFilter`.

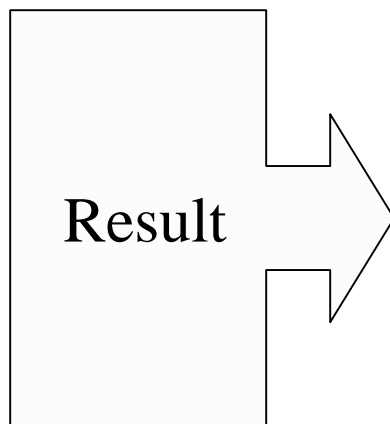
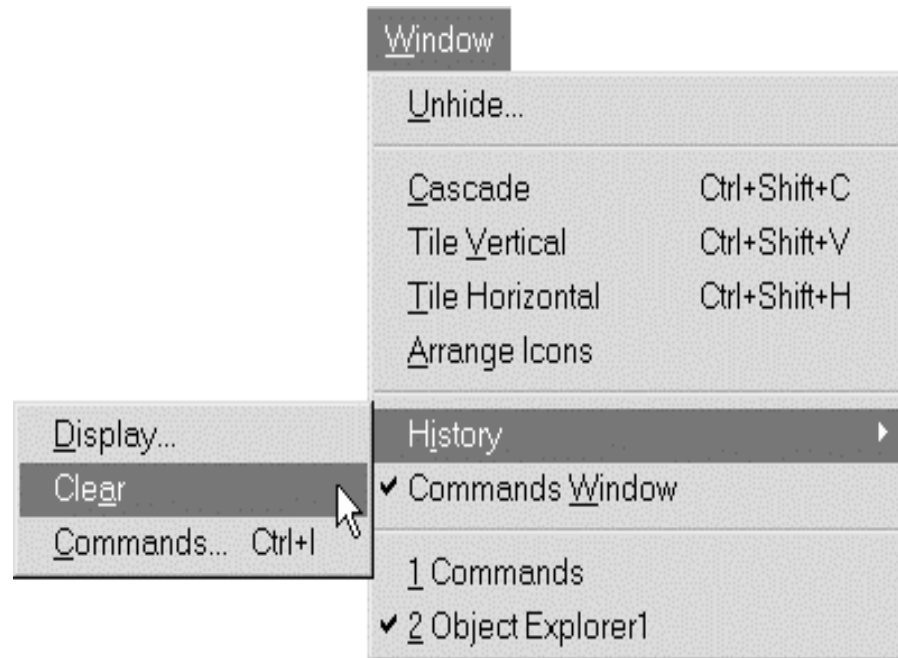
```
1 1
guiOpen( "ObjectBrowser",
    FileName = "d:\\Splus45\\users\\fhilal\\_Prefs\\Defa
)
Hide = F,
Show = "Normal",
Top = "Auto",
Left = "Auto",
Width = "420",
Height = "430")

guiCreate( "BrowserPage", Name = "Object Browser$Page 1
    GroupItems = "",
    BrowserTabToolTip = "",
    ImageFileName = "",
    ShowOnlyFolders = F,
    FilterString = "(All)",
    CaseSensitiveFilter = T,
    DatabaseFilter = "",
```

- Close the window.
- Choose No if prompted to save the **History Log**.

# Clearing the History Log

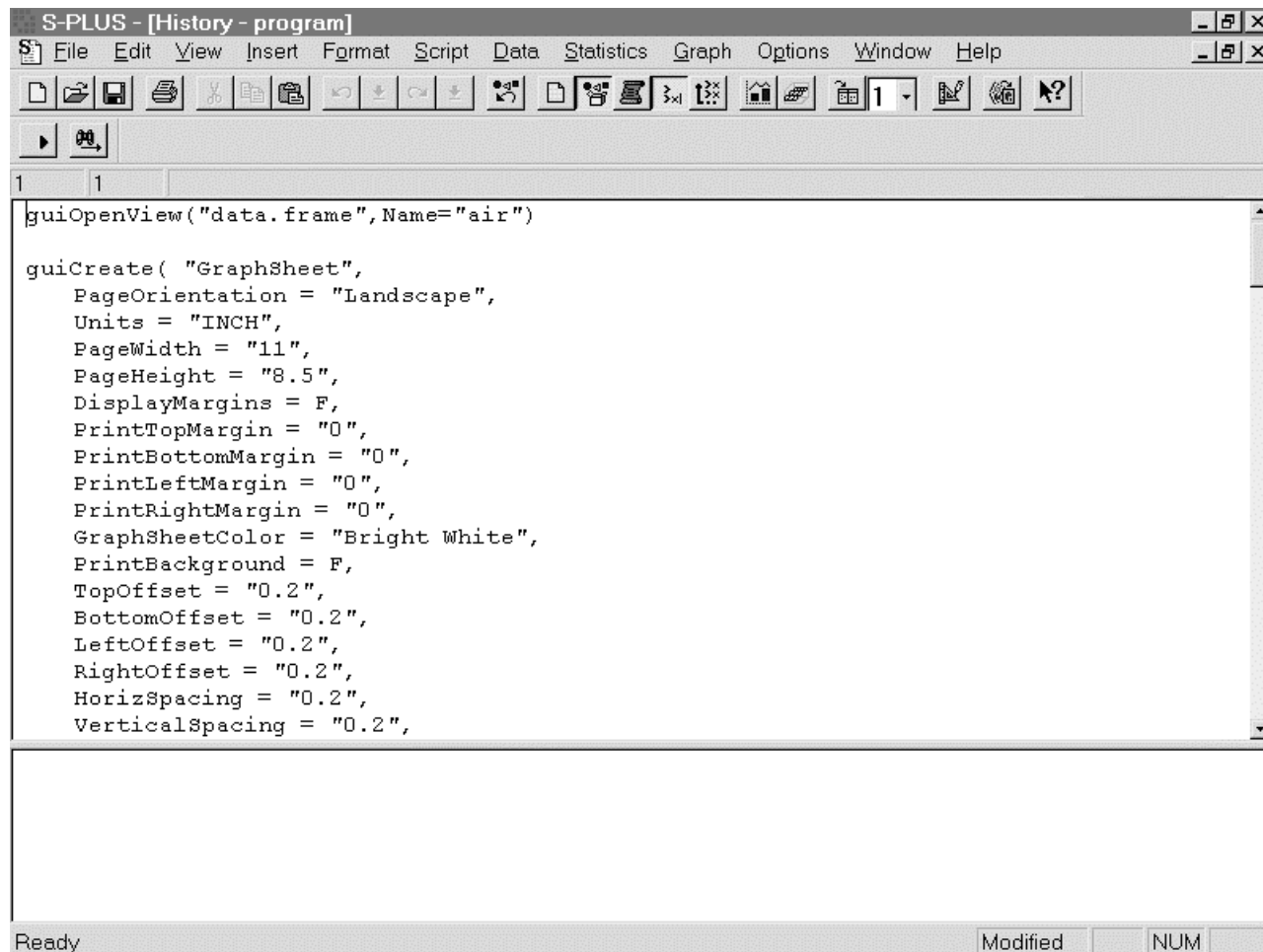
- Click **Window**
- Click **History**
- Click **Clear**



There is **NO History Log**.

# Creating the History Log

- For the “**air**” data frame
  - open the data window
  - create a scatter plot for radiation versus ozone
  - Insert a title “ Scatter Plot”
  - Save the Graph as “Graph1.sgr”
- Now look at the **History Log**.



# Creating Script Files from the History Log

- Open the “**History Log**” window
- Do “**Save As**” from the “**File**” menu
- Specify the name of the file (The extension is “ssc”). Name the file **Mygraph.ssc**
- Hit Save

## Editing Script Files

- Scroll down to the statement you want to change
- Change it
- Save the file

## Example: Editing Script Files

- Change the Main Title from “ Scatter Plot” to “Scatter Plot of Radiation versus Ozone”

- Scroll down to the statement you want to change

```
guiCreate( "MainTitle", Name = "GS1$1$1",  
UseDate = F,  
UseTime = F,  
Title = "Scatter Plot",  
xPosition = "Auto",
```

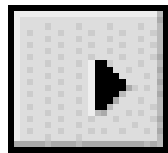
- Change

```
– Title = "Scatter Plot", To Title = "Scatter Plot of  
Radiation versus Ozone"
```

- Save the file

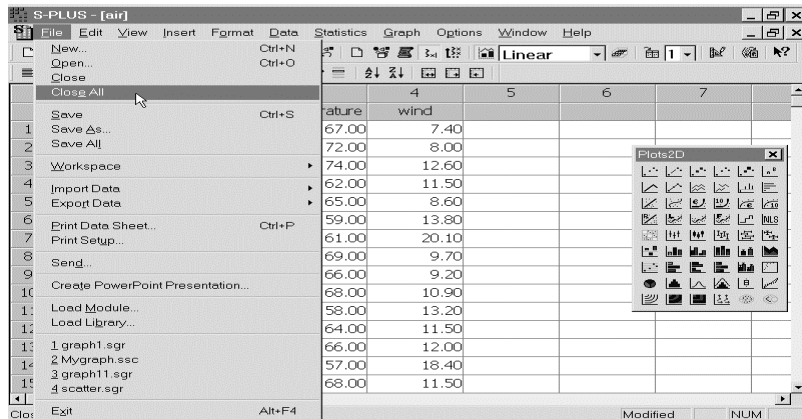
## Executing Script Files

- From the “**File**” Menu choose “**Close All**”
- From the “**File**” Menu in the recently opened files list select “MyGraph.ssc”
- Click on the “**Run**” button to run the script file

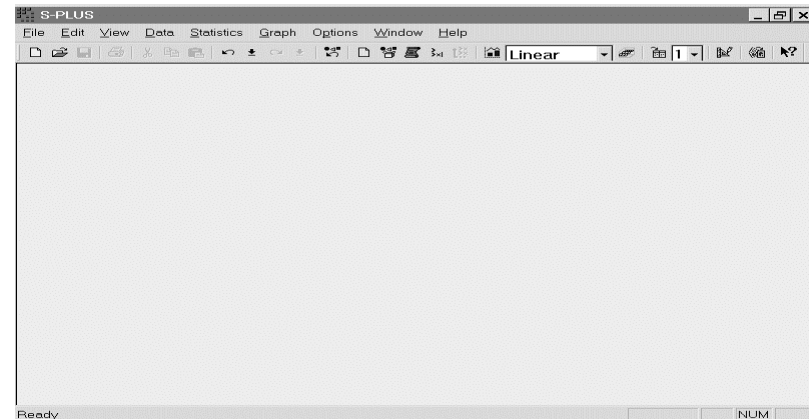


# Example: Executing the Script Files

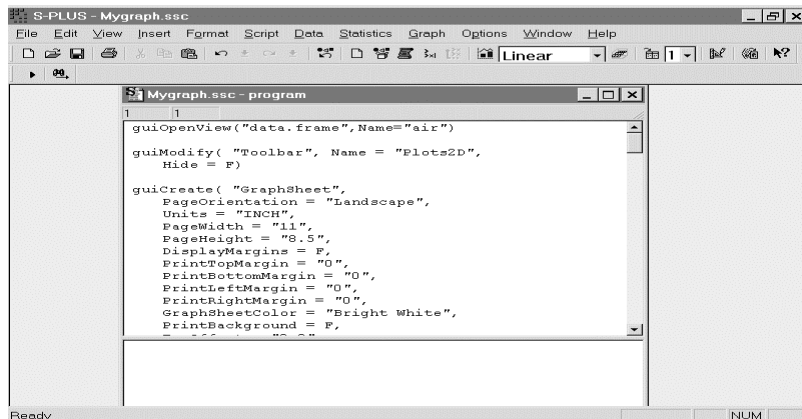
Step 1



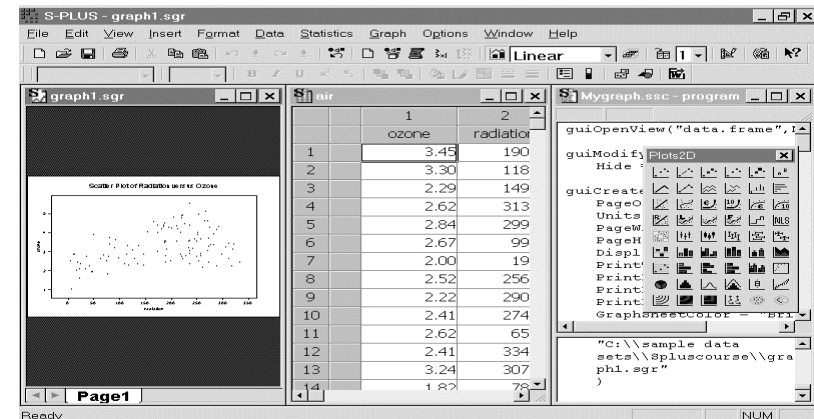
Step 2



Step 3



Result



# Linking the History Log to a Toolbar Button

- To run this session automatically by clicking a button, we need to attach the history log to a toolbar button.

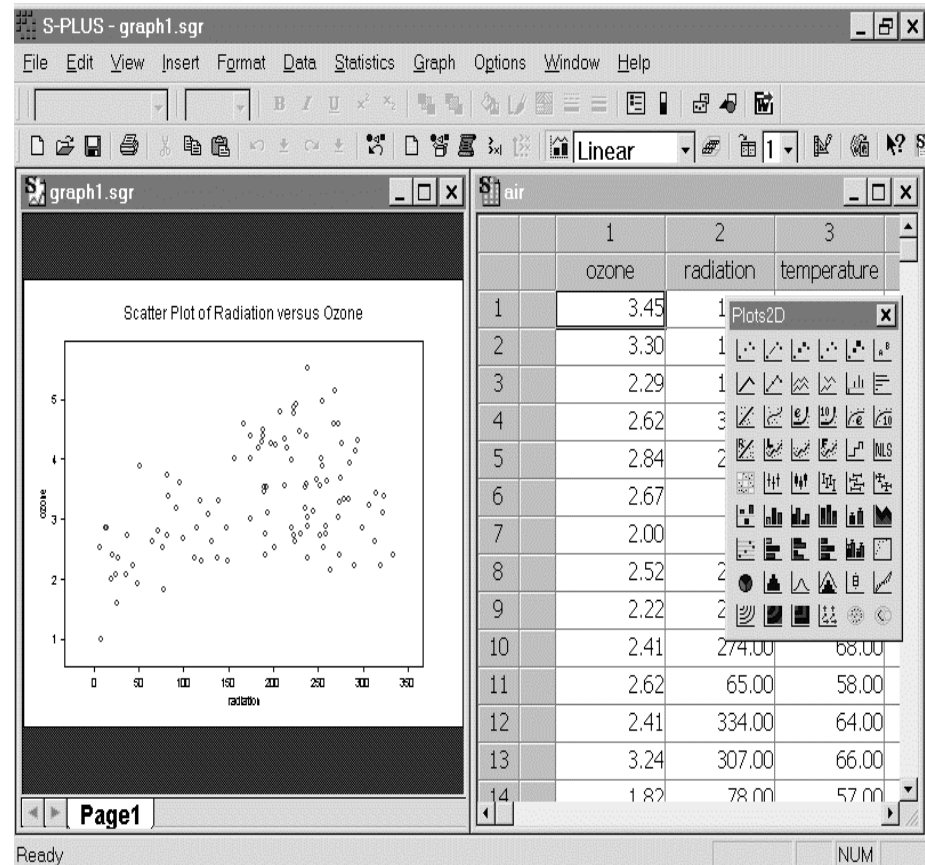
- Open the “**History Log**”
- Select the desired commands in the History Log by dragging the mouse until just after the following commands

```
guiSave( "Script", Name = "History",  
        NewName = "mygraph",  
        FileName = "C:\\sample data sets\\mygraph.ssc")  
guiSave( "Script", Name = "mygraph",  
        FileName = "C:\\sample data sets\\mygraph.ssc")
```

- Link the History Log to a button by dragging the selected command onto the main toolbar and dropping them anywhere in the toolbar



- To test the button
  - From the File menu, choose
    - Close All
    - Click No when prompt to save any file
    - Click on the new toolbar button

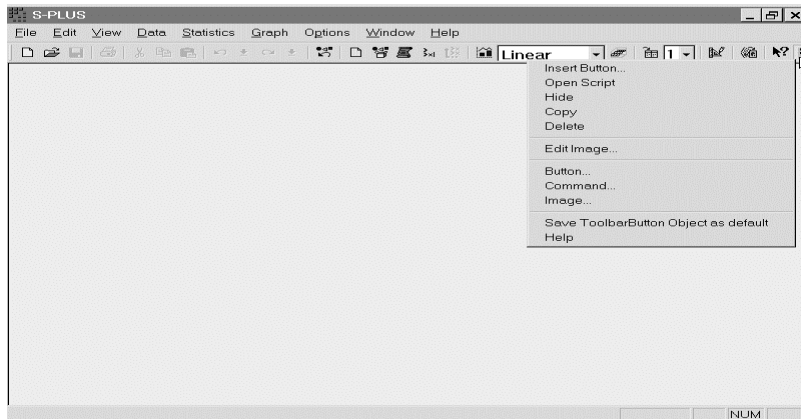


# Modifying Toolbar buttons

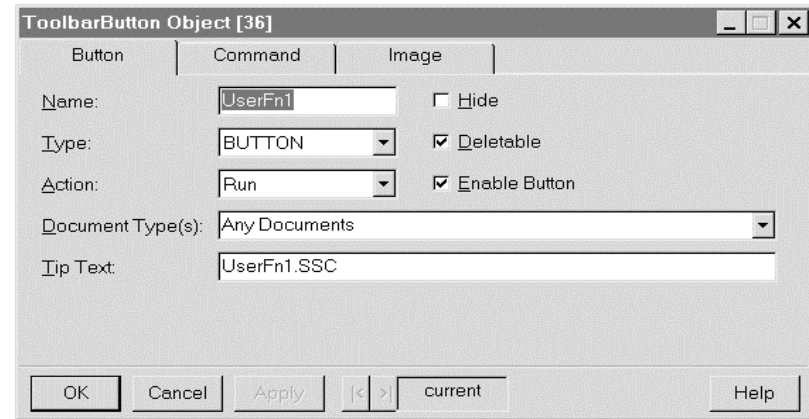
- To edit the tooltip text
  - Right-Click on the button to display a context menu
  - Choose “**Button**” from the menu
  - In the Tip Text field, type “**My Graph**”
  - Click Ok
- To reposition the toolbar button
  - Hold down the ALT key while dragging the toolbar button to a new location
  - Release the mouse to drop the button

# Example: Edit the tooltip text

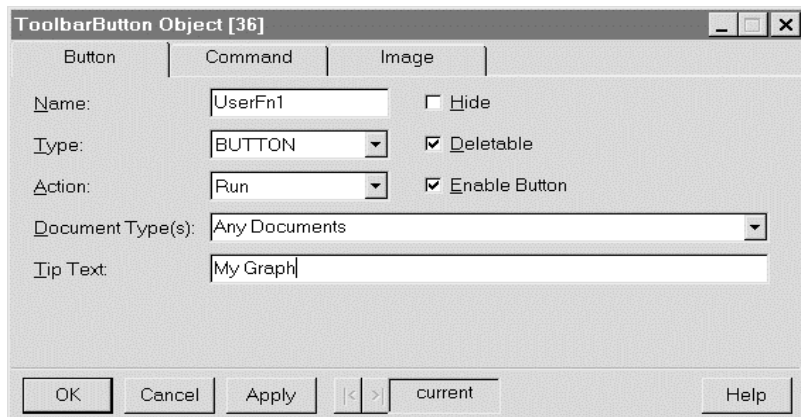
Step 1



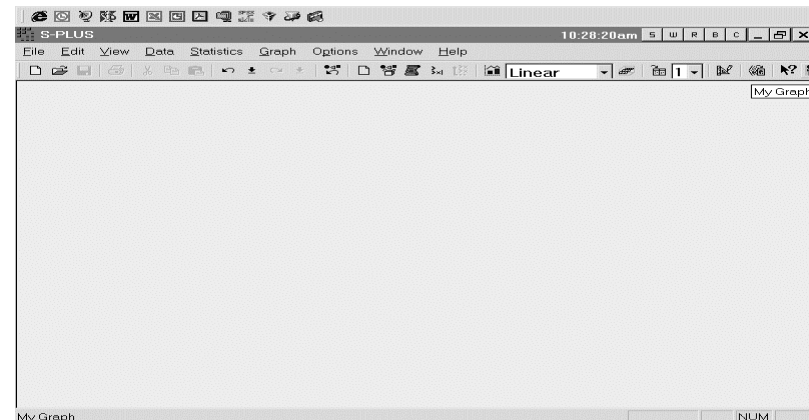
Step 2



Step 3



Result



# EXERCISES

# DATA Import & Manipulation

1. Import the following data sets
  - a. c:\Spluscourse\gpa.txt
  - b. c:\Spluscourse\cardata.xls
  - c. c:\Spluscourse\gpadata.sd2
2. Import c:\Spluscourse\cardata.xls by choosing only the small type cars. This can be done by specifying the following statement in the filter page **where Type = Small** . Call the data frame car.small.

### 3. For the **gpa** data frame

- a. Create a new variable called  
“School.Average” which equals to  
“(HSM+HSS+HSE)/3”
- b. Change the column names as follows  
HSM To HighS.Math.Avg  
HSS To HighS.Science.Avg  
HSE To HighS.English.Avg
- c. Create two data sets one that contains the  
Female data and the other that contains the  
Males data.

- d. Create a subset of the **gpa** data frame, extracting all observations for which the GPA is less than the mean of all of the GPA. (The `mean(var)` automatically computes the mean of `var`)
4. Merge the **authors** and **books** data frames as in the examples, except click the **Data Frame 1** box in the **Include Non-Matched Rows in** section of the dialogue. What happened? How does it differ from the outcome in the examples?

5. Merge the **authors** and **books** data frames as in the examples, except click the **Data Frame 2** box in the **Include Non-Matched Rows in** section of the dialogue. What happened? How does it differ from the outcome in the examples?

# DATA Export

6. Export the following data frames

a. **gas** to a SAS data set

b. **gun** to an ASCII File

7. For the **GPA** data frame.

Create a Histogram for each of the variables.

Create a powerpoint presentation for those graphs.

# Graphs

8. For the GPADATA do the following
  - a. Create a 2D Scatter Plot with SATM in the X-axis and SATV in the y-axis
    - Change the x-axis Label to “SAT MATH SCORE”
    - Change the y-axis Label to “SAT VERBAL SCORE”
    - Change the Font Size for the two Axis Title to 20.
    - the Color for the two Axis Title to whatever you like.
    - Let the x-axis and y-axis range be from 0 to 800 by 100.
    - Rotate the x-axis label to 30 degrees

**b.** Create a Histogram Plot for one of the variables then,

Insert a title

Change the fill color, fill pattern, pattern color, border color and the border weight.

**c.** Place the two plots on one Graph sheet.

**d.** Send the Graph to a Power Point Presentation.

**9.** For the **gpadata** data frame create a histogram for the SATM where  $SATM > 600$

# Summary Statistics

**10.** For the **gpadata** data frame do the following

- a.** Compute Summary Statistics for both variables
- b.** Construct a Linear Regression model. Take SATM as you independent variable
- c.** Create a 2D- scatter plot with the regression line

# History Log & Script Files

11.

- Clear you History Log
- For the **gpa** data frame do the following
  - Create a histogram for the SATM
  - Insert a title
- Open the history log
- Create a Script File from the history log
- Change your Title and Save the Script File Again

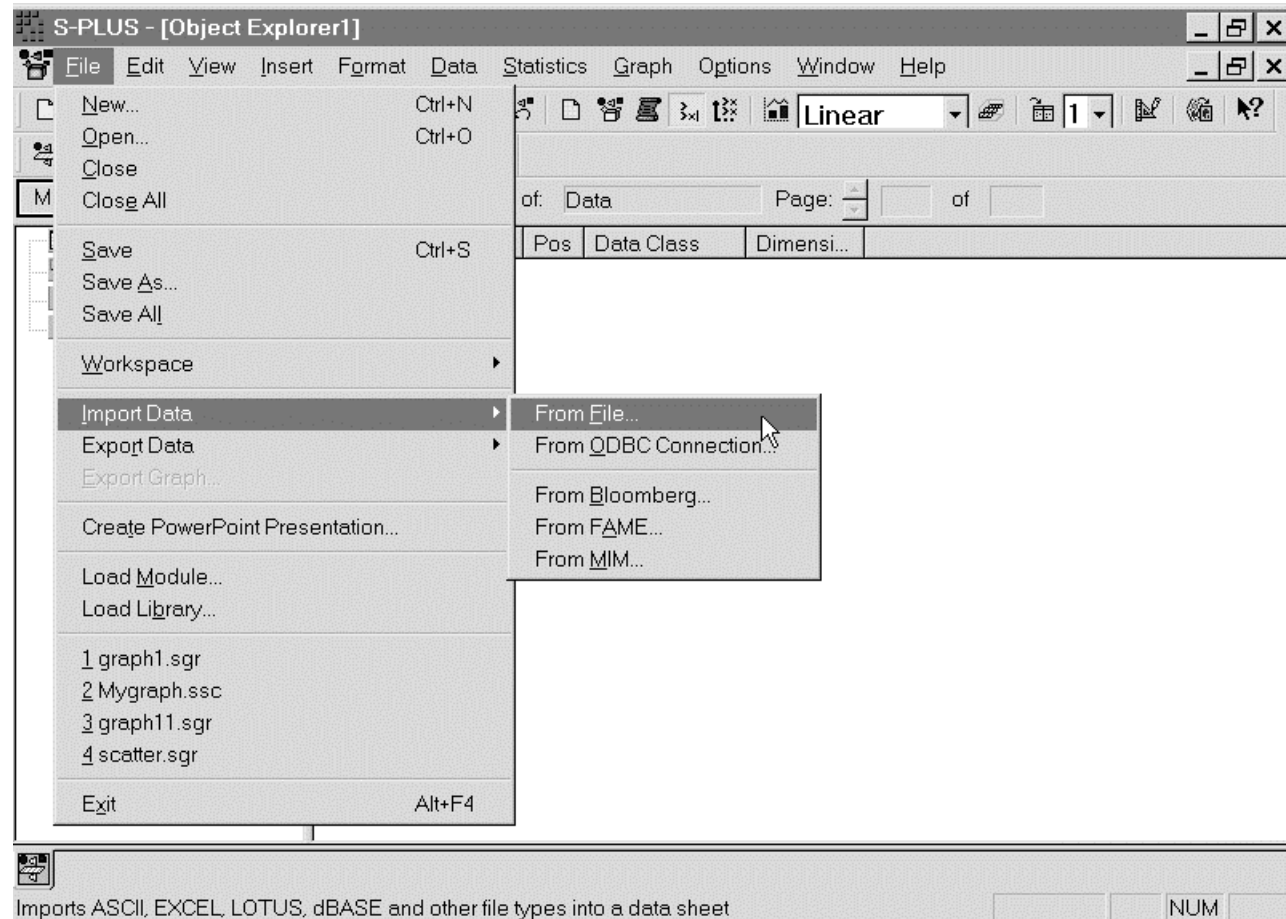
- Execute the script file you saved
- Link the history log to a toolbar button
- Change the tooltip text to something meaningful



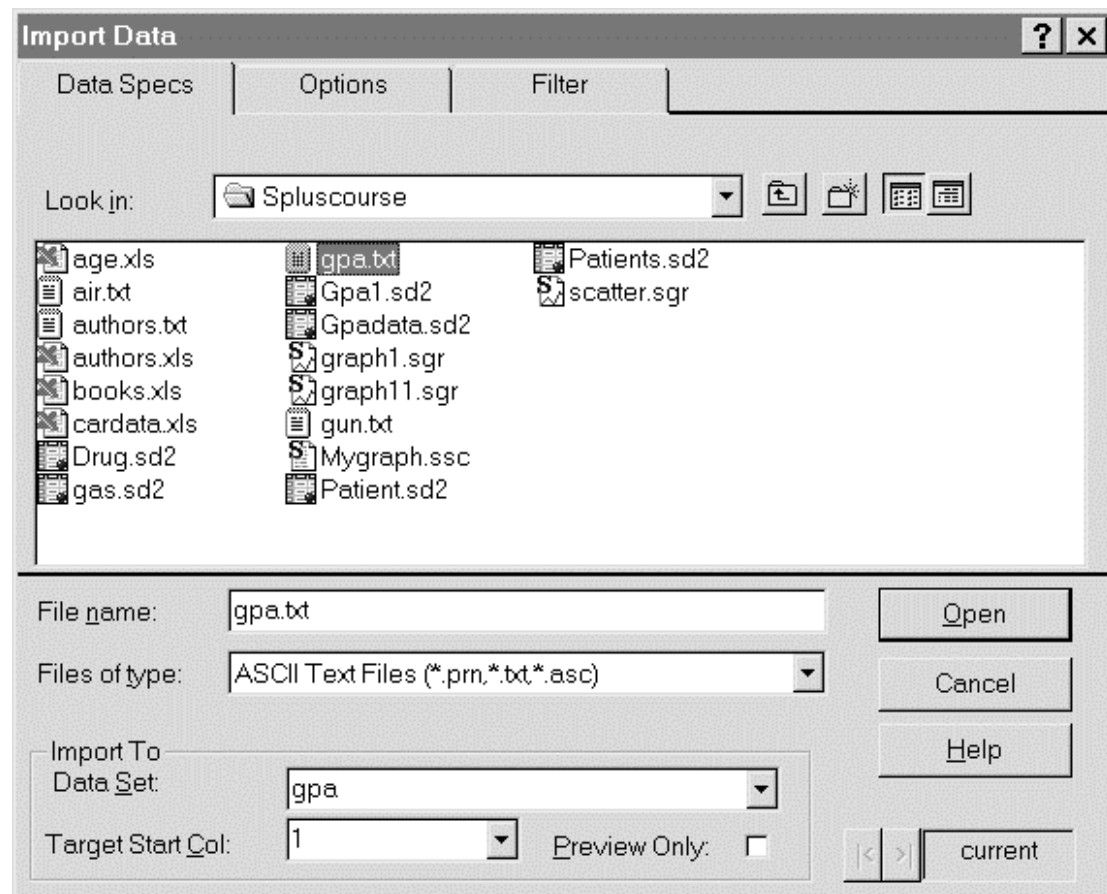
# Solution

## 1. a. Import c:\Spluscourse\gpa.txt

**Step 1**



## Step 2



**Result**

S-PLUS - [gpa]

File Edit View Insert Format Data Statistics Graph Options Window Help

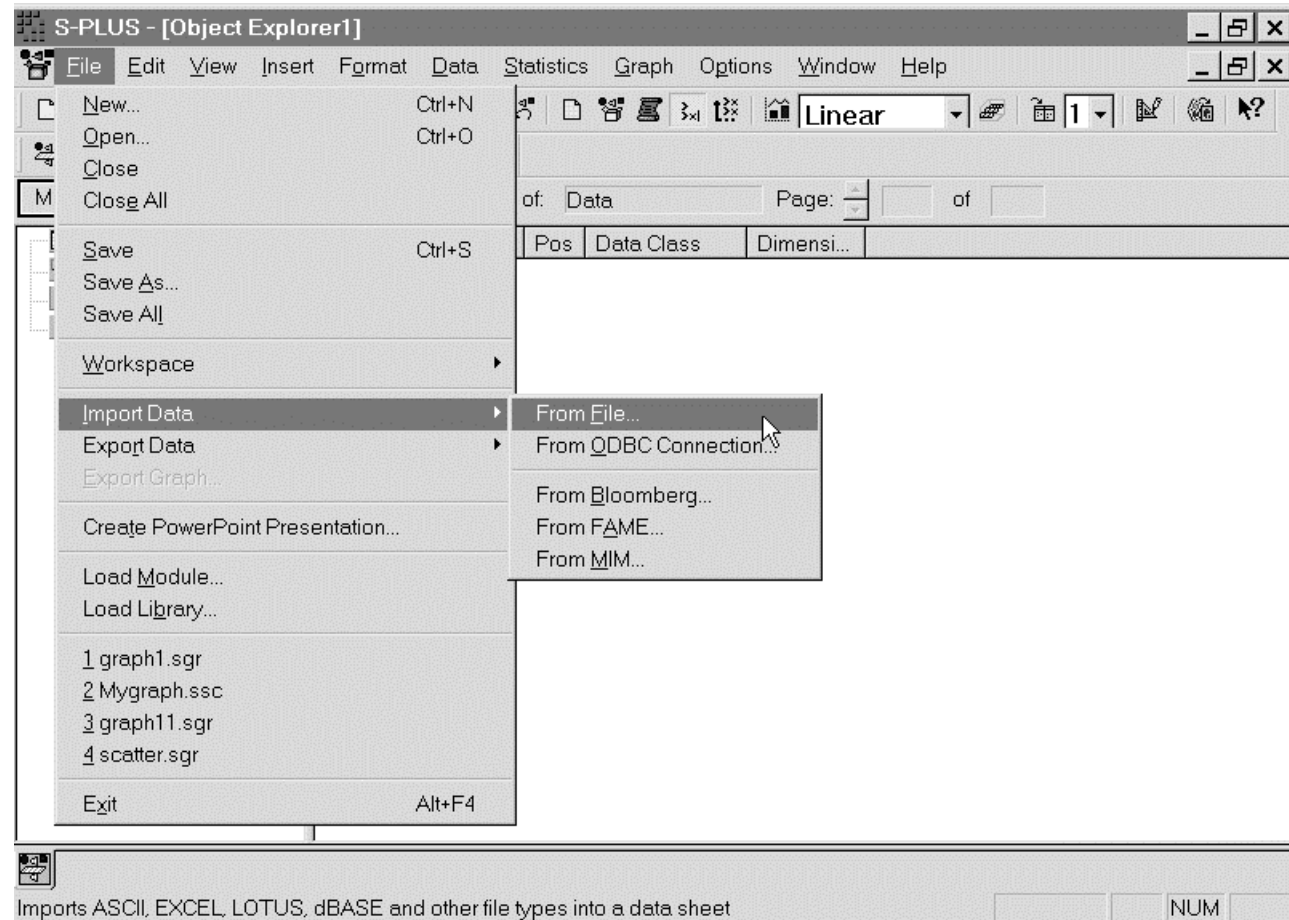
Linear 1

	1	2	3	4	5	6	7
	GPA	HSM	HSS	HSE	SEX		
1	5.32	10.00	10.00	10.00	Female		
2	5.14	9.00	9.00	10.00	Male		
3	3.84	9.00	6.00	6.00	Female		
4	5.34	10.00	9.00	9.00	Male		
5	4.26	6.00	8.00	5.00	Female		
6	4.35	8.00	6.00	8.00	Female		
7	5.33	9.00	7.00	9.00	Male		
8	4.85	10.00	8.00	8.00	Male		
9	4.76	10.00	10.00	10.00	Male		
10	5.72	7.00	8.00	7.00	Female		
11	4.08	9.00	10.00	7.00	Female		
12	5.38	8.00	9.00	8.00	Female		
13	2.40	6.00	6.00	7.00	Female		
14	5.50	8.00	7.00	8.00	Female		
15	5.69	10.00	10.00	8.00	Male		

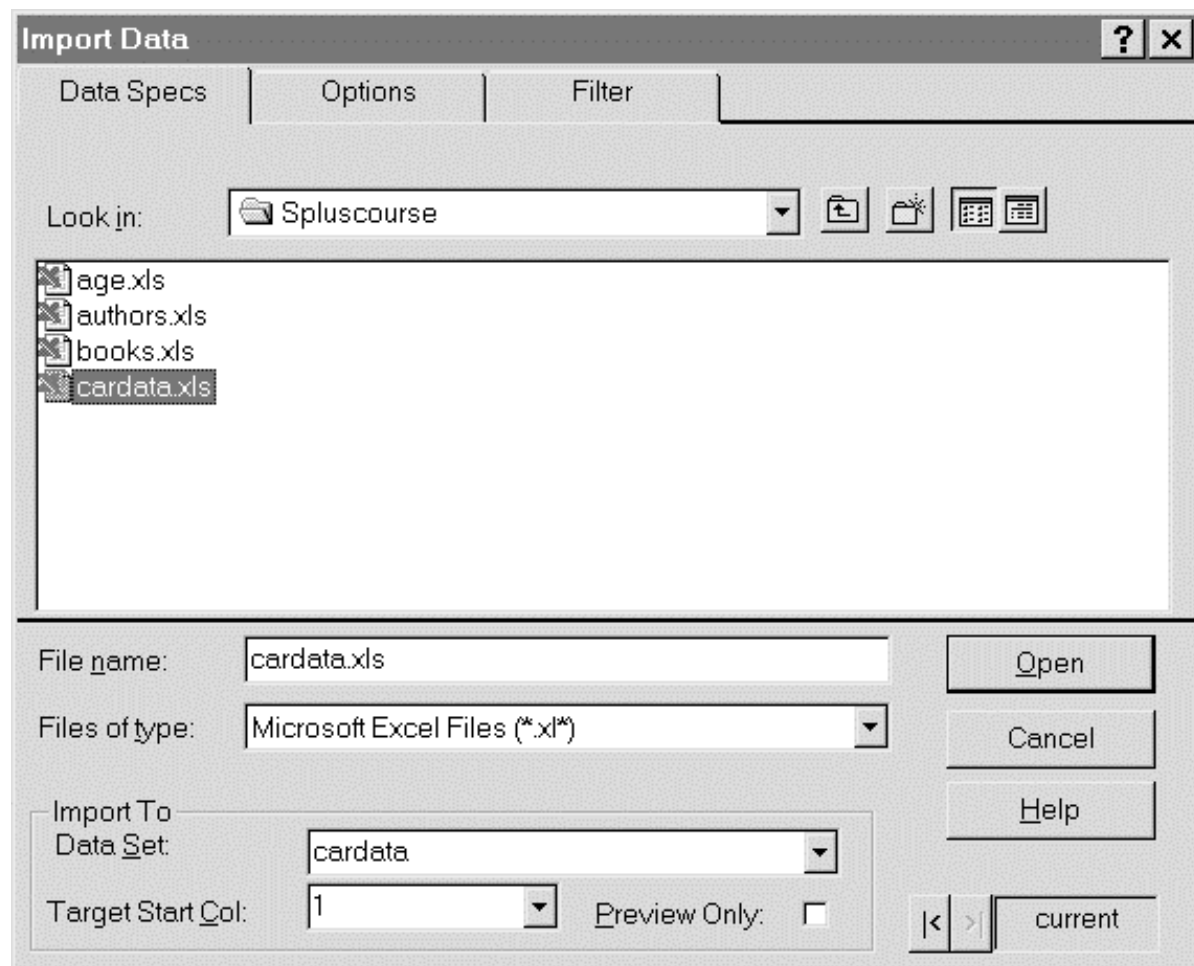
Ready NUM

## 1. b. Import c:\Spluscourse\cardata.xls

**Step 1**



## Step 2



**Rresult**

S-PLUS - [cardata]

File Edit View Insert Format Data Statistics Graph Options Window Help

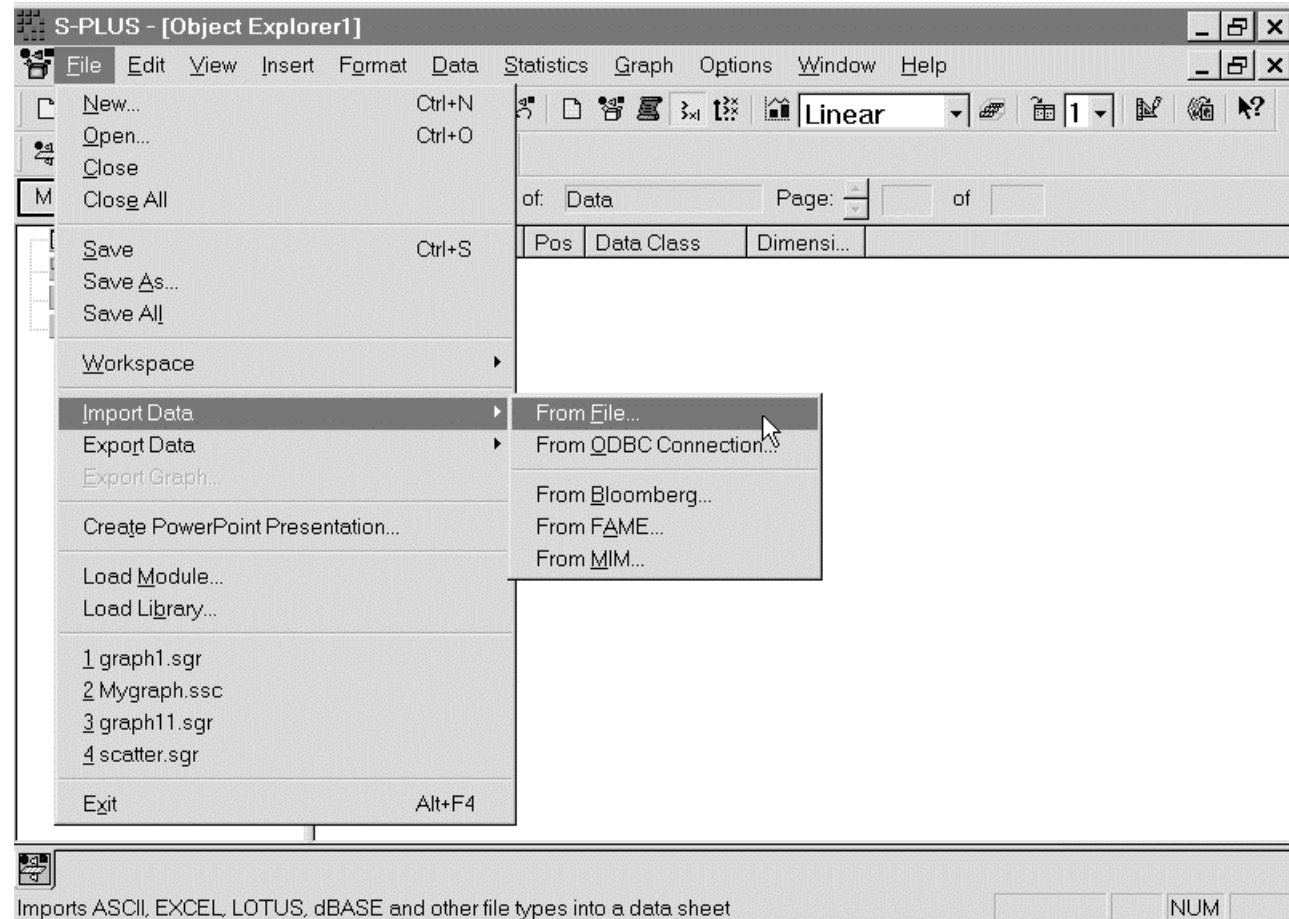
Linear 1

	1	2	3	4	5	6	7	
	Price	Country	Reliability	Mileage	Type	Weight	Disp.	
1	8895.00	USA	4	33.00	Small	2560.00	97.00	
2	7402.00	USA	2	33.00	Small	2345.00	114.00	
3	6319.00	Korea	4	37.00	Small	1845.00	81.00	
4	6635.00	Japan/USA	5	32.00	Small	2260.00	91.00	
5	6599.00	Japan	5	32.00	Small	2440.00	113.00	
6	8672.00	Mexico	4	26.00	Small	2285.00	97.00	
7	7399.00	Japan/USA	5	33.00	Small	2275.00	97.00	
8	7254.00	Korea	1	28.00	Small	2350.00	98.00	
9	9599.00	Japan	5	25.00	Small	2295.00	109.00	
10	5866.00	Japan	NA	34.00	Small	1900.00	73.00	
11	8748.00	Japan/USA	5	29.00	Small	2390.00	97.00	
12	6488.00	Japan	5	35.00	Small	2075.00	89.00	
13	9995.00	Germany	3	26.00	Small	2330.00	109.00	
14	11545.00	USA	1	20.00	Sporty	3320.00	305.00	
15	9745.00	USA	1	27.00	Sporty	2885.00	153.00	

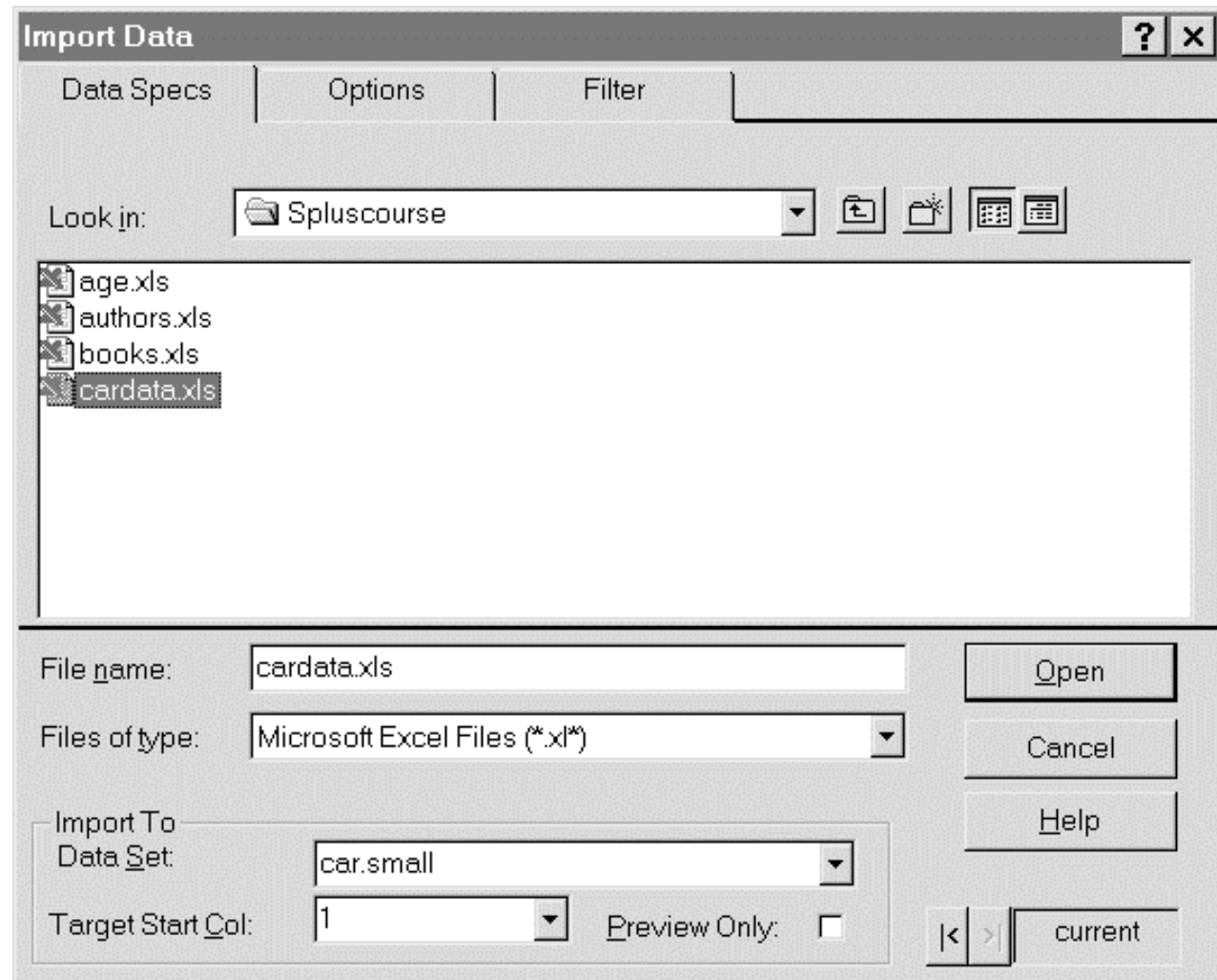
Ready NUM

## 2. Import c:\Spluscourse\cardata.xls where Type = Small

Step 1



## Step 2a



## Step 2b

**Import Data** ? x

Data Specs | Options | Filter

Filter Information:

where Type = Small

File name: cardata.xls

Files of type: Microsoft Excel Files (\*.xls)

Import To Data Set: car.small

Target Start Col: 1

Preview Only: ☐

Open Cancel Help

|< >| current

**Result**

S-PLUS - [car.small]

File Edit View Insert Format Data Statistics Graph Options Window Help

Linear 1

	1	2	3	4	5	6	7	
	Price	Country	Reliability	Mileage	Type	Weight	Disp.	
1	8895.00	USA	4	33.00	Small	2560.00	97.00	
2	7402.00	USA	2	33.00	Small	2345.00	114.00	
3	6319.00	Korea	4	37.00	Small	1845.00	81.00	
4	6635.00	Japan/USA	5	32.00	Small	2260.00	91.00	
5	6599.00	Japan	5	32.00	Small	2440.00	113.00	
6	8672.00	Mexico	4	26.00	Small	2285.00	97.00	
7	7399.00	Japan/USA	5	33.00	Small	2275.00	97.00	
8	7254.00	Korea	1	28.00	Small	2350.00	98.00	
9	9599.00	Japan	5	25.00	Small	2295.00	109.00	
10	5866.00	Japan	NA	34.00	Small	1900.00	73.00	
11	8748.00	Japan/USA	5	29.00	Small	2390.00	97.00	
12	6488.00	Japan	5	35.00	Small	2075.00	89.00	
13	9995.00	Germany	3	26.00	Small	2330.00	109.00	
14								
15								

Ready NUM

3. a.

Step 1

S-PLUS - [gpa]

File Edit View Insert Format Data Statistics Graph Options Window Help

Linear 1

	1	2	3	4	5	6	7
	GPA	HSM	HSS	HSE	SEX		
1	5.32	10.00	10.00	10.00	Female		
2	5.14	9.00	9.00	10.00	Male		
3	3.84	9.00	6.00	6.00	Female		
4	5.34	10.00	9.00	9.00	Male		
5	4.26	6.00	8.00	5.00	Female		
6	4.35	8.00	6.00	8.00	Female		
7	5.33	9.00	7.00	9.00	Male		
8	4.85	10.00	8.00	8.00	Male		
9	4.76	10.00	10.00	10.00	Male		
10	5.72	7.00	8.00	7.00	Female		
11	4.08	9.00	10.00	7.00	Female		
12	5.38	8.00	9.00	8.00	Female		
13	2.40	6.00	6.00	7.00	Female		
14	5.50	8.00	7.00	8.00	Female		
15	5.69	10.00	10.00	8.00	Male		

Ready NUM

3. a.

**Step 2**

**Insert Columns**

Name(s)

Fill Expression:

Start Column:  Count:

Data Set:  Column Type:

OK Cancel Apply < > current Help

3. a.

**Result**

S-PLUS - [gpa]

File Edit View Insert Format Data Statistics Graph Options Window Help

Linear 1

	1	2	3	4	5	6	7
	GPA	HSM	HSS	HSE	SEX	School.Average	
1	5.32	10.00	10.00	10.00	Female	10.00	
2	5.14	9.00	9.00	10.00	Male	9.33	
3	3.84	9.00	6.00	6.00	Female	7.00	
4	5.34	10.00	9.00	9.00	Male	9.33	
5	4.26	6.00	8.00	5.00	Female	6.33	
6	4.35	8.00	6.00	8.00	Female	7.33	
7	5.33	9.00	7.00	9.00	Male	8.33	
8	4.85	10.00	8.00	8.00	Male	8.67	
9	4.76	10.00	10.00	10.00	Male	10.00	
10	5.72	7.00	8.00	7.00	Female	7.33	
11	4.08	9.00	10.00	7.00	Female	8.67	
12	5.38	8.00	9.00	8.00	Female	8.33	
13	2.40	6.00	6.00	7.00	Female	6.33	
14	5.50	8.00	7.00	8.00	Female	7.67	
15	5.69	10.00	10.00	8.00	Male	9.33	

Ready Modified NUM

3. b.

Step 1

S-PLUS - [gpa]

File Edit View Insert Format Data Statistics Graph Options Window Help

Linear 1

	1	2	3	4	5	6	7
	GPA	HSM				School.Average	
1	5.32	1				10.00	
2	5.14					9.33	
3	3.84					7.00	
4	5.34					9.33	
5	4.26					6.33	
6	4.35					7.33	
7	5.33					8.33	
8	4.85					8.67	
9	4.76					10.00	
10	5.72					7.33	
11	4.08					8.67	
12	5.38	8.00	9.00	8.00	Female	8.33	
13	2.40	6.00	6.00	7.00	Female	6.33	
14	5.50	8.00	7.00	8.00	Female	7.67	
15	5.69	10.00	10.00	8.00	Male	9.33	

Modified NUM

3. b.

**Step 2**

**Double Precision Column [2]**

Name: HighS.Math.Avg      Format Type: Decimal

Width: 12      Precision: 2

Justification: Right

Description:

OK   Cancel   Apply   |< >|   current   Help

3. b.

**Result**

S-PLUS - [gpa]

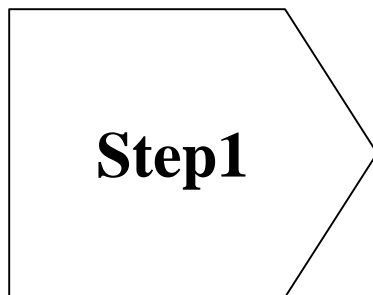
File Edit View Insert Format Data Statistics Graph Options Window Help

Linear 1

	1	2	3	4	5	6
	GPA	HighS.Math.Avg	HighS.Science.Avg	HighS.English.Avg	SEX	School.Average
1	5.32	10.00	10.00	10.00	Female	10.00
2	5.14	9.00	9.00	10.00	Male	9.33
3	3.84	9.00	6.00	6.00	Female	7.00
4	5.34	10.00	9.00	9.00	Male	9.33
5	4.26	6.00	8.00	5.00	Female	6.33
6	4.35	8.00	6.00	8.00	Female	7.33
7	5.33	9.00	7.00	9.00	Male	8.33
8	4.85	10.00	8.00	8.00	Male	8.67
9	4.76	10.00	10.00	10.00	Male	10.00
10	5.72	7.00	8.00	7.00	Female	7.33
11	4.08	9.00	10.00	7.00	Female	8.67
12	5.38	8.00	9.00	8.00	Female	8.33
13	2.40	6.00	6.00	7.00	Female	6.33
14	5.50	8.00	7.00	8.00	Female	7.67
15	5.69	10.00	10.00	8.00	Male	9.33

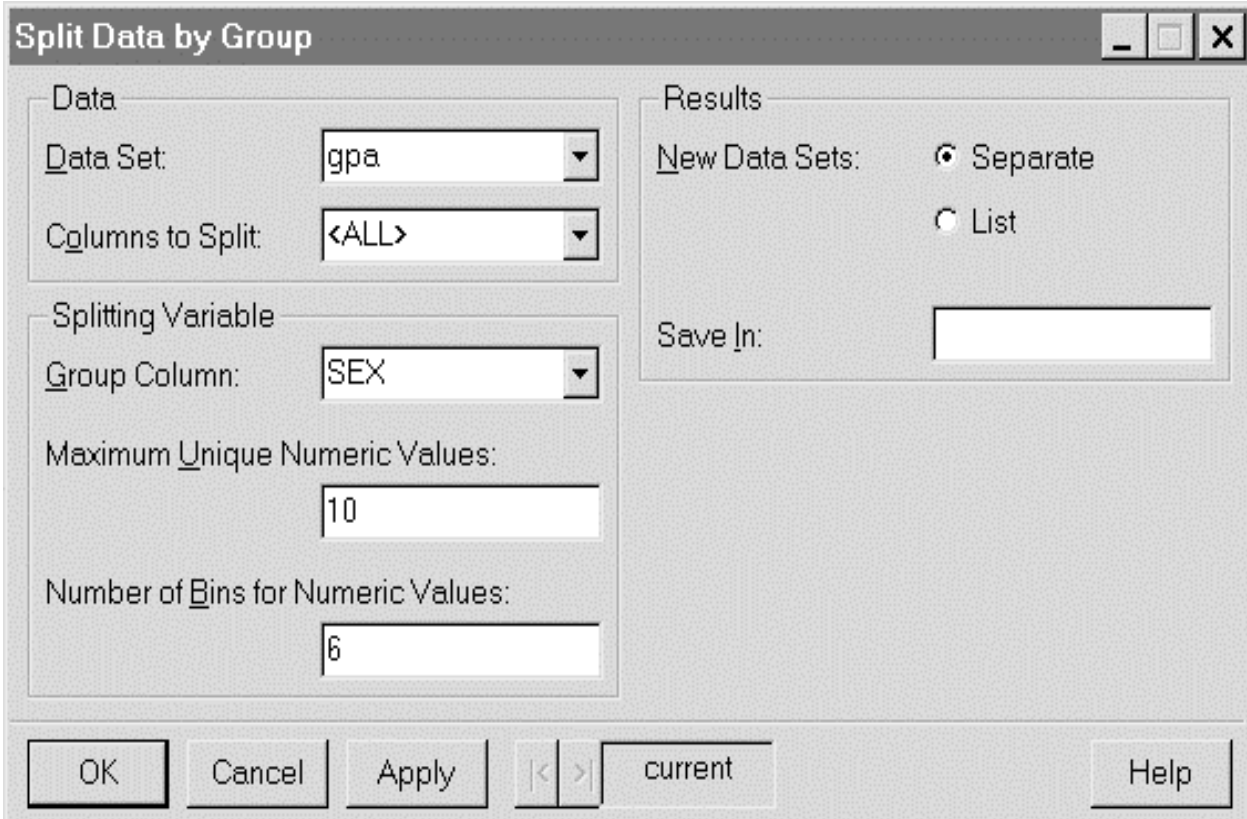
Ready NUM

**3. c.**



3. c.

**Step2**



The image shows the 'Split Data by Group' dialog box in SPSS. The 'Data' section has 'Data Set' set to 'gpa' and 'Columns to Split' set to '<ALL>'. The 'Splitting Variable' section has 'Group Column' set to 'SEX'. The 'Maximum Unique Numeric Values' is set to 10, and the 'Number of Bins for Numeric Values' is set to 6. The 'Results' section has 'New Data Sets' set to 'Separate' (selected with a radio button) and 'List' (unselected). The 'Save In:' field is empty. At the bottom, there are buttons for 'OK', 'Cancel', 'Apply', '<' '>', 'current', and 'Help'.

**Split Data by Group**

**Data**

Data Set: gpa

Columns to Split: <ALL>

**Splitting Variable**

Group Column: SEX

Maximum Unique Numeric Values: 10

Number of Bins for Numeric Values: 6

**Results**

New Data Sets: ☒ Separate ☐ List

Save In:

OK Cancel Apply < > current Help

3. c.

**Result**

S-PLUS - gpa.Male

File Edit View Insert Format Data Statistics Graph Options Window Help

Linear 1

gpa.Male

		1	2	3	4	5	6	7
		GPA	ighS.Math.Avg	ghS.Science.Avg	ghS.English.A	SEX	chool.Averag	
1	2	5.14	9.00	9.00	10.00	Male	9.33	
2	4	5.34	10.00	9.00	9.00	Male	9.33	
3	7	5.33	9.00	7.00	9.00	Male	8.33	
4	8	4.85	10.00	8.00	8.00	Male	8.67	
5	9	4.75	10.00	10.00	10.00	Male	10.00	

gpa.Female

		1	2	3	4	5	6	7
		GPA	ighS.Math.Avg	ghS.Science.Avg	ghS.English.A	SEX	chool.Averag	
1	1	5.32	10.00	10.00	10.00	Female	10.00	
2	3	3.84	9.00	6.00	6.00	Female	7.00	
3	5	4.26	6.00	8.00	5.00	Female	6.33	
4	6	4.35	8.00	6.00	8.00	Female	7.33	
5	10	5.72	7.00	8.00	7.00	Female	7.33	
6	11	4.00	8.00	10.00	7.00	Female	8.00	

Ready NUM

3. d.

Step 1

The screenshot shows the S-PLUS software interface with a data table and a menu open. The menu is titled 'Select Data...' and contains the following options: Copy, Move, Transpose, Exchange, Remove, Clear, Restructure, Fill..., Recode..., Transform..., Create Categories..., Random Numbers..., Distribution Functions..., Split..., Subset... (highlighted by a mouse cursor), Merge..., Tabulate..., Expand Grid..., Random Sample..., and Change Data Type....

	1	2	4	5	6
	GPA	HighS.Math	HighS.English.Avg	SEX	School.Average
1	5.32	1	10.00	Female	10.00
2	5.14		10.00	Male	9.33
3	3.84		6.00	Female	7.00
4	5.34	1	9.00	Male	9.33
5	4.26		5.00	Female	6.33
6	4.35		8.00	Female	7.33
7	5.33		9.00	Male	8.33
8	4.85	1	8.00	Male	8.67
9	4.76	1	10.00	Male	10.00
10	5.72		7.00	Female	7.33
11	4.08		7.00	Female	8.67
12	5.38		8.00	Female	8.33
13	2.40		7.00	Female	6.33
14	5.50		8.00	Female	7.67
15	5.69	1	8.00	Male	9.33

3. d.

**Step 2**

**Subset**

**Data**

Data Set: gpa

Columns in Subset: <ALL>

**Subset Expression**

Subset Rows with: GPA < mean(GPA)

Create Subset

**Results**

Result Type: ☒ Data Set ☐ Add Column

Save In: low.gpa

Target Column: Subset

OK Cancel Apply |< >| current Help

3. d.

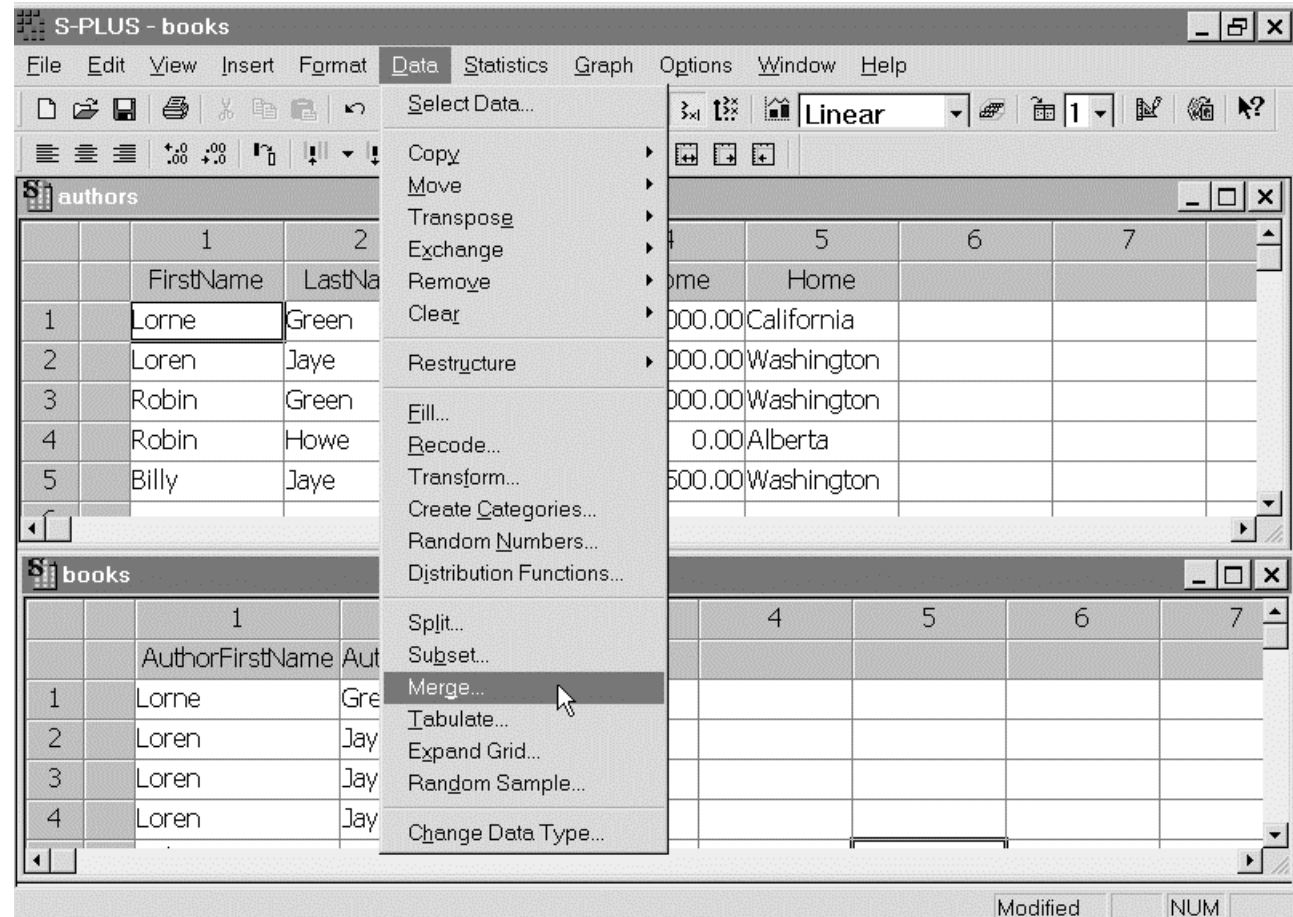
**Result**

S-PLUS - [low.gpa]							
File Edit View Insert Format Data Statistics Graph Options Window Help							
Linear 1							
		1	2	3	4	5	6
		GPA	HighS.Math.Avg	HighS.Science.Avg	HighS.English.Avg	SEX	chool.Avera
1	3	3.84	9.00	6.00	6.00	Female	7.00
2	5	4.26	6.00	8.00	5.00	Female	6.3
3	6	4.35	8.00	6.00	8.00	Female	7.3
4	11	4.08	9.00	10.00	7.00	Female	8.6
5	13	2.40	6.00	6.00	7.00	Female	6.3
6	18	3.80	7.00	7.00	7.00	Male	7.00
7	20	4.57	9.00	10.00	10.00	Male	9.6
8	23	4.28	8.00	10.00	10.00	Male	9.3
9	26	3.60	4.00	7.00	7.00	Male	6.00
10	28	4.00	3.00	7.00	6.00	Female	5.3
11	31	4.38	9.00	9.00	10.00	Male	9.3
12	32	4.58	10.00	10.00	9.00	Male	9.6
13	33	4.34	7.00	7.00	6.00	Female	6.6
14	34	4.26	5.00	7.00	7.00	Female	6.3
15	36	4.03	6.00	7.00	9.00	Female	7.3

ReadyModifiedNUM

4.

Step 1



4.

**Step 2**

**Merge Two Data Sets**

Data

Data Set 1: books

Data Set 2: authors

Row Matching

Match by:

- ☐ All Common Cols
- ☐ Row Names
- ☒ Specified Cols

Columns in Set 1: AuthorFirstName

Columns in Set 2: FirstName, LastN

Include Non-Matched Rows in

☒ Data Set 1

☐ Data Set 2

Suffix for Non-Matching Common Cols

Data Set 1: .1

Data Set 2: .2

Results

Save In: books.4

OK Cancel Apply < > current Help

4.

**Result**

S-PLUS - [books.4]

File Edit View Insert Format Data Statistics Graph Options Window Help

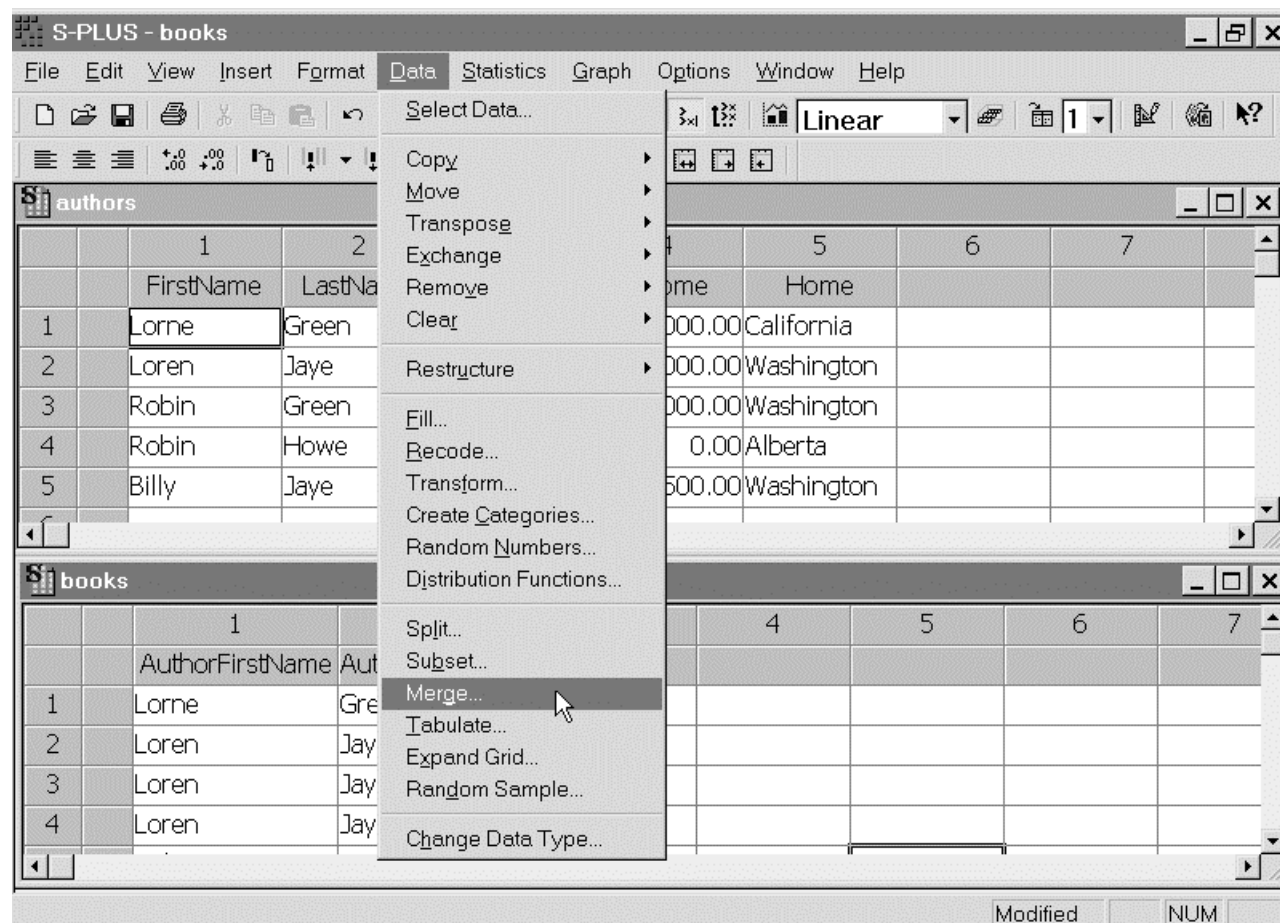
Linear 1

	1	2	3	4	5	6	7
	AuthorFirstNam	AuthorLastNam	Book	Age	Income	Home	
1	Loren	Jaye	Midwifery	40.00	40000.00	Washington	
2	Loren	Jaye	Gardening	40.00	40000.00	Washington	
3	Loren	Jaye	Perennials	40.00	40000.00	Washington	
4	Lorne	Green	Bonanza	82.00	1200000.00	California	
5	Rich	Calaway	Splus	NA	NANA		
6	Robin	Green	Who_dun_it?	45.00	25000.00	Washington	
7							
8							
9							
10							
11							
12							
13							
14							
15							

Ready NUM

5.

Step 1



5.

## Step 2

**Merge Two Data Sets**

Data

Data Set 1: books

Data Set 2: authors

Row Matching

Match by:

- ☐ All Common Cols
- ☐ Row Names
- ☒ Specified Cols

Columns in Set 1: AuthorFirstName,

Columns in Set 2: FirstName, LastN

Include Non-Matched Rows in

☐ Data Set 1

☒ Data Set 2

Suffix for Non-Matching Common Cols

Data Set 1: .1

Data Set 2: .2

Results

Save In: books.5

OK Cancel Apply |< >| current Help

5.

Result

S-PLUS - [books.5]

File Edit View Insert Format Data Statistics Graph Options Window Help

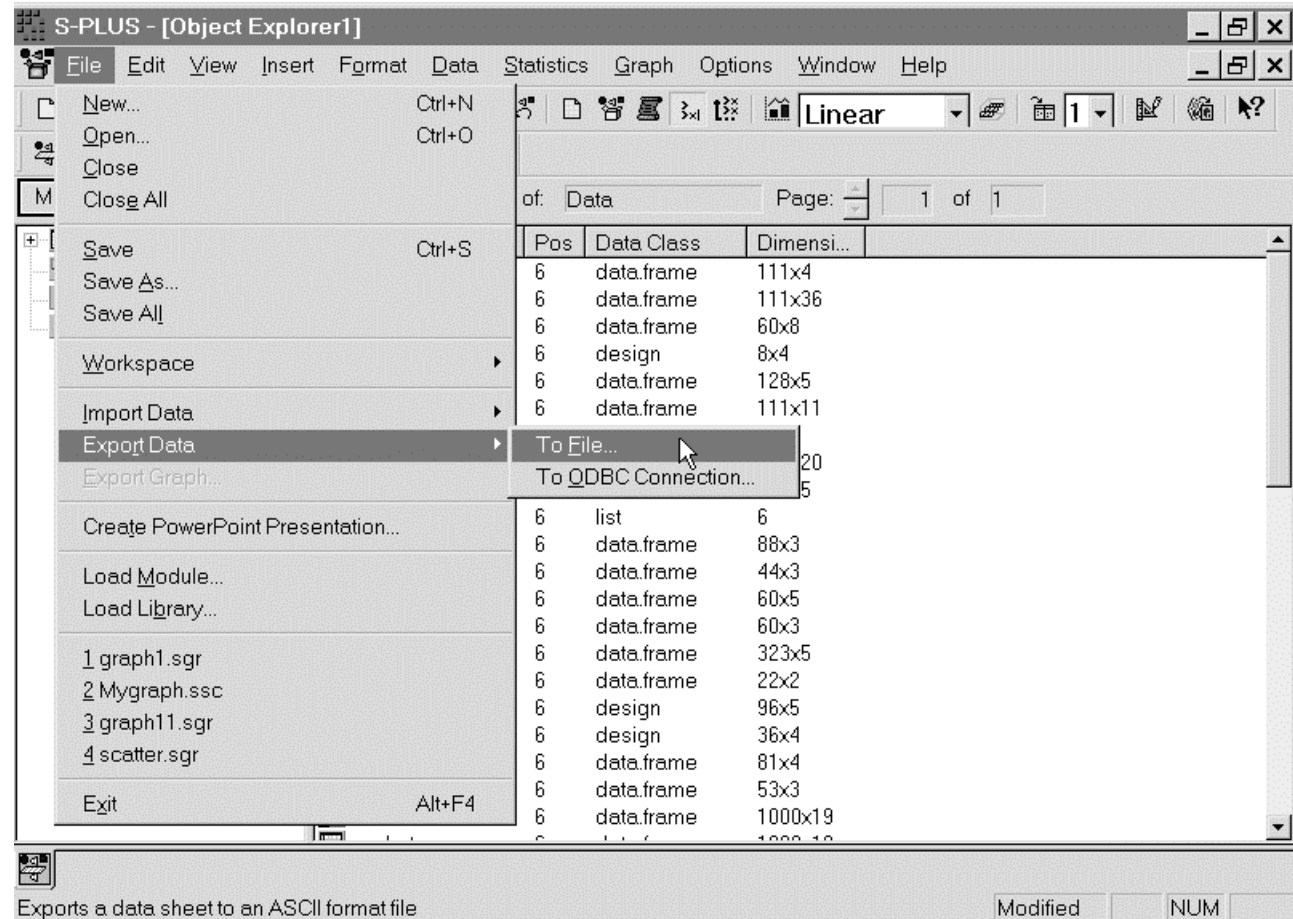
Linear 1

	1	2	3	4	5	6	7
	uthorFirstNam	uthorLastNam	Book	Age	Income	Home	
1	Billy	Jaye	NA	40.00	27500.00	Washington	
2	Loren	Jaye	Midwifery	40.00	40000.00	Washington	
3	Loren	Jaye	Gardening	40.00	40000.00	Washington	
4	Loren	Jaye	Perennials	40.00	40000.00	Washington	
5	Lorne	Green	Bonanza	82.00	1200000.00	California	
6	Robin	Green	Who_dun_it?	45.00	25000.00	Washington	
7	Robin	Howe	NA	2.00	0.00	Alberta	
8							
9							
10							
11							
12							
13							
14							
15							

Ready NUM

6. a.

Step 1



6. a.

**Step 2**

**Export Data** [?] [X]

Data Specs | **Options**

Save in: Spluscourse [v] [Folder Icon] [New Folder Icon] [Grid Icon] [List Icon]

- Drug.sd2
- Gpa1.sd2
- Gpadata.sd2
- Patient.sd2
- Patients.sd2

File name: gas.sd2 [v]

Save as type: SAS Files (\*.sd2) [v]

Export From Data Set: gas [v]

Column(s): ALL [v] Row(s): ALL [v]

[Save] [Cancel] [Help]

6. a.

**Result**

SAS - [VIEWTABLE: TMP1.gas]

File Edit Data View Customize Globals Options Window Help

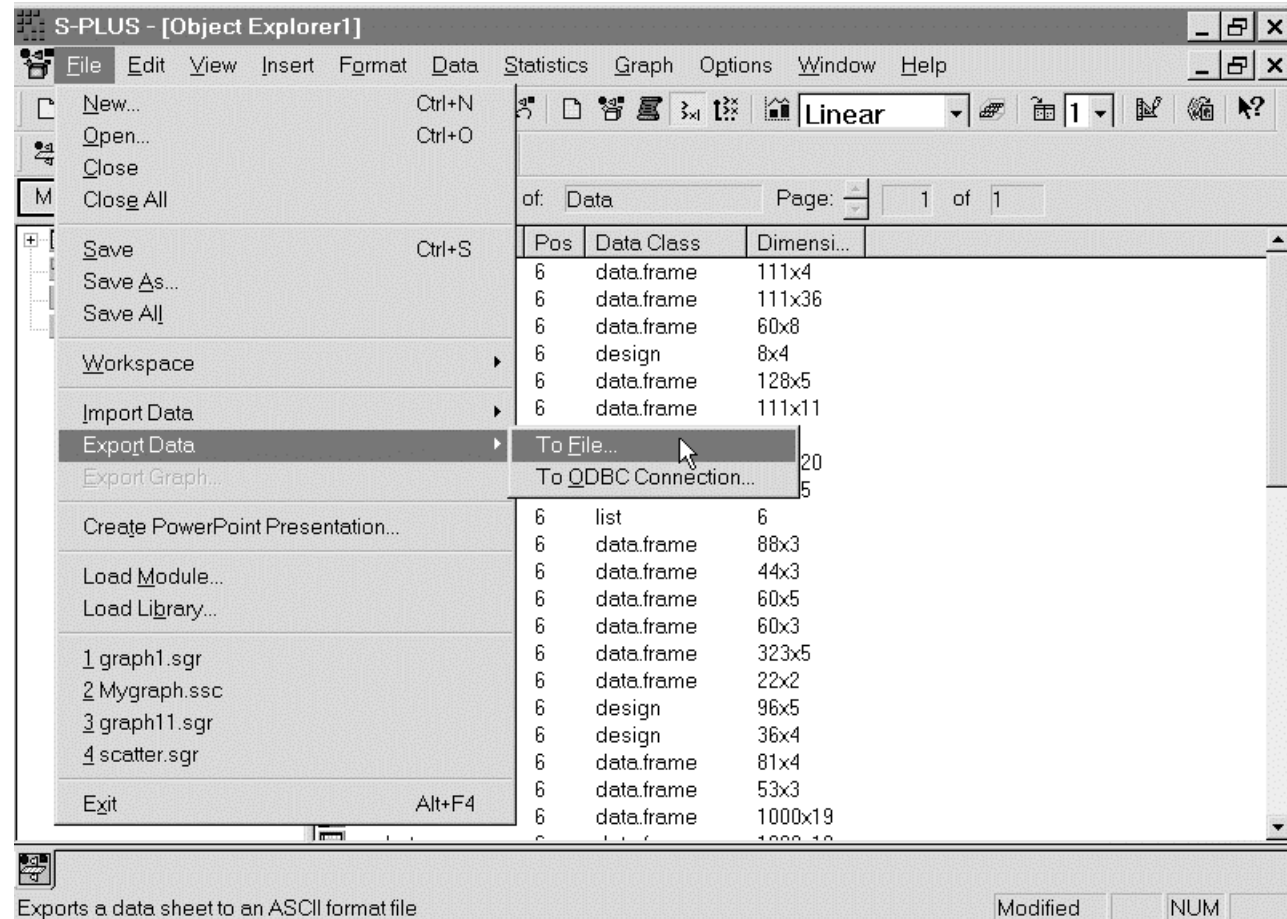
	ROWNAMES	NOX	E
1	1	4.818	0.831
2	2	2.849	1.045
3	3	3.275	1.021
4	4	4.691	0.97
5	5	4.255	0.825
6	6	5.064	0.891
7	7	2.118	0.71
8	8	4.602	0.801
9	9	2.286	1.074
10	10	0.97	1.148
11	11	3.965	1
12	12	5.344	0.928
13	13	3.834	0.767
14	14	1.99	0.701
15	15	5.199	0.807
16	16	5.283	0.902
17	17	3.752	0.997
18	18	0.537	1.224
19	19	1.64	1.089
20	20	5.055	0.973
21	21	4.937	0.98
22	22	1.551	0.555

NOTE: Table has been opened with ROW level edit access.

C:\WINDOWS\system32

6. b.

Step 1



6. b.

Step 2

**Export Data** [?] [X]

Data Specs | **Options**

Save in: Spluscourse [v] [Folder Icon] [New Folder Icon] [Grid Icon] [Table Icon]

- air.txt
- authors.txt
- gpa.txt

File name: gun.txt

Save as type: ASCII Text Files (\*.txt\*.prn) [v]

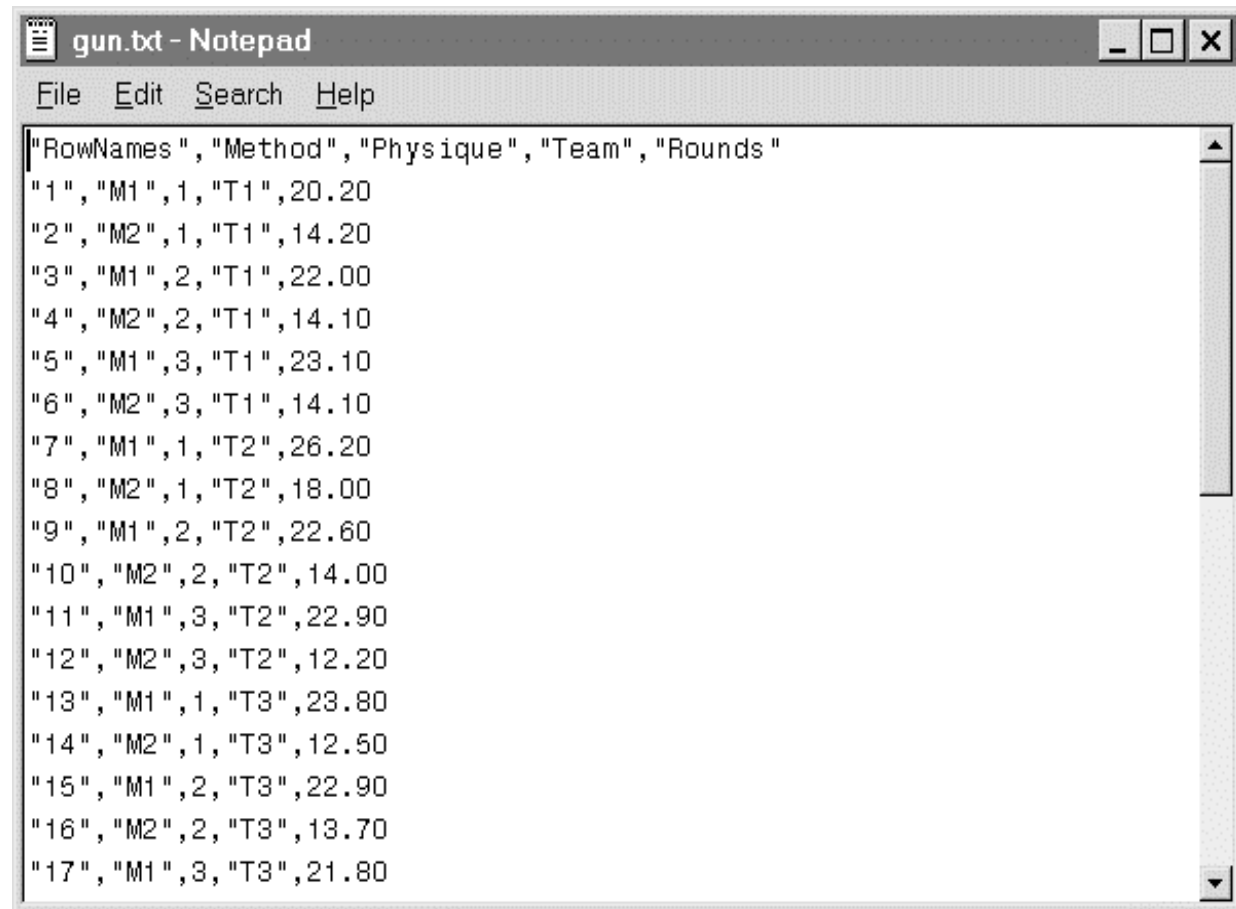
Export From Data Set: gun [v]

Column(s): ALL [v] Row(s): ALL [v]

[Save] [Cancel] [Help]

6. b.

**Result**



```
gun.txt - Notepad
File Edit Search Help
"RowNames", "Method", "Physique", "Team", "Rounds "
"1", "M1", 1, "T1", 20.20
"2", "M2", 1, "T1", 14.20
"3", "M1", 2, "T1", 22.00
"4", "M2", 2, "T1", 14.10
"5", "M1", 3, "T1", 23.10
"6", "M2", 3, "T1", 14.10
"7", "M1", 1, "T2", 26.20
"8", "M2", 1, "T2", 18.00
"9", "M1", 2, "T2", 22.60
"10", "M2", 2, "T2", 14.00
"11", "M1", 3, "T2", 22.90
"12", "M2", 3, "T2", 12.20
"13", "M1", 1, "T3", 23.80
"14", "M2", 1, "T3", 12.50
"15", "M1", 2, "T3", 22.90
"16", "M2", 2, "T3", 13.70
"17", "M1", 3, "T3", 21.80
```

9.

**Step 1**

**Histogram/Density [1]**

**Data to Plot** | Options | Histogram Bars | Density Line

**Data Columns**

Data Set: Gpadata

x Columns: SATM

y Columns:

z Columns:

w Columns:

**Override Conditioning**

Type: Auto

Data Set: Gpadata

Columns:

Draw in Panels: All

**Scale to**

X Axis #: 1

Y Axis #: 1

Subset Rows with: SATM > 600

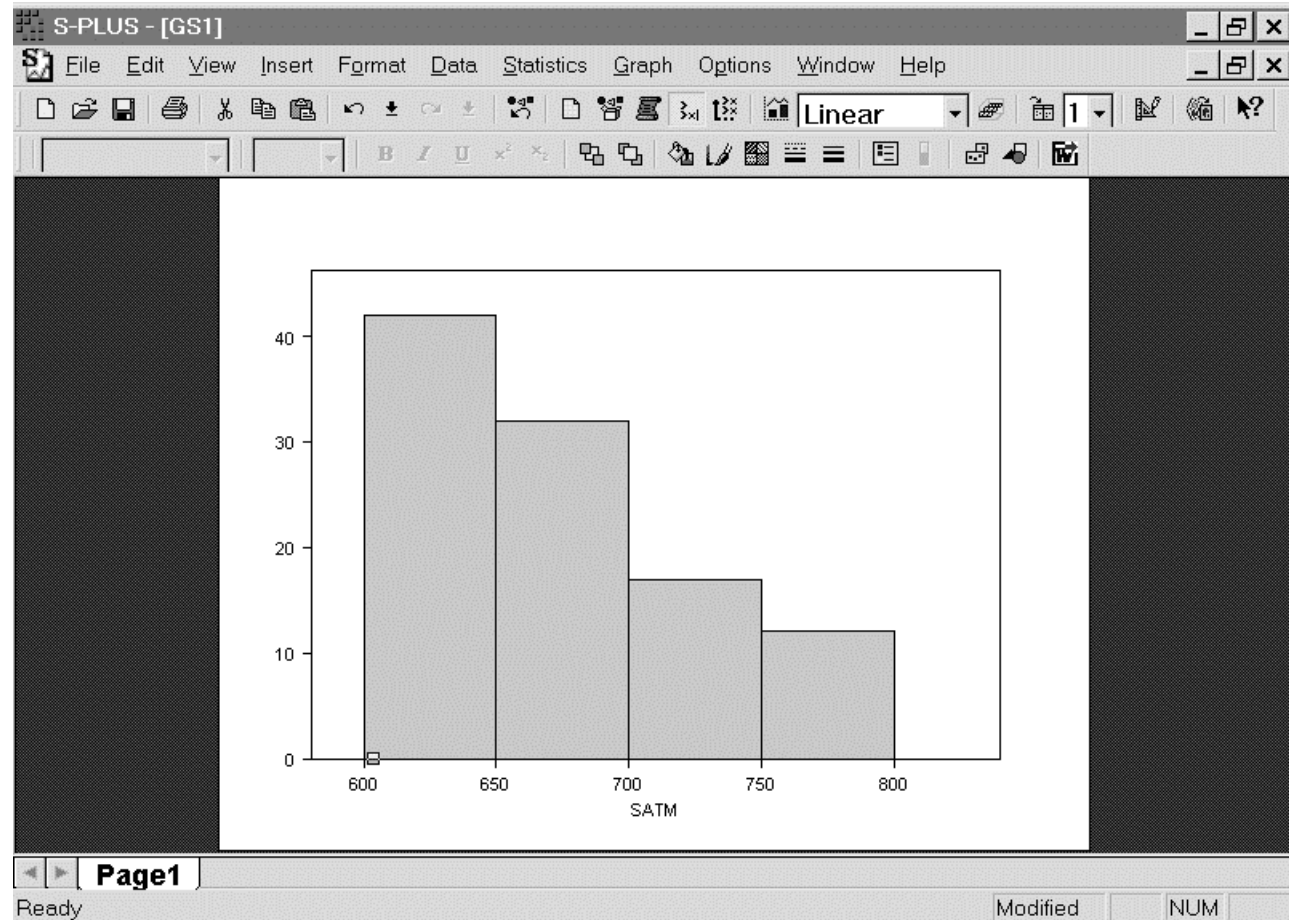
☐ Hide

☒ Crop

OK Cancel Apply |< >| current Help

9.

**Result**



10. a.

Step 1

The screenshot shows the S-PLUS software interface with a data table and a statistics menu open. The data table has two columns, 1 (SATM) and 2 (SATV), with 15 rows of data. The value 530.00 in row 4, column 2 is highlighted. The statistics menu is open, showing options like Summary Statistics, Crosstabulations, and Correlations.

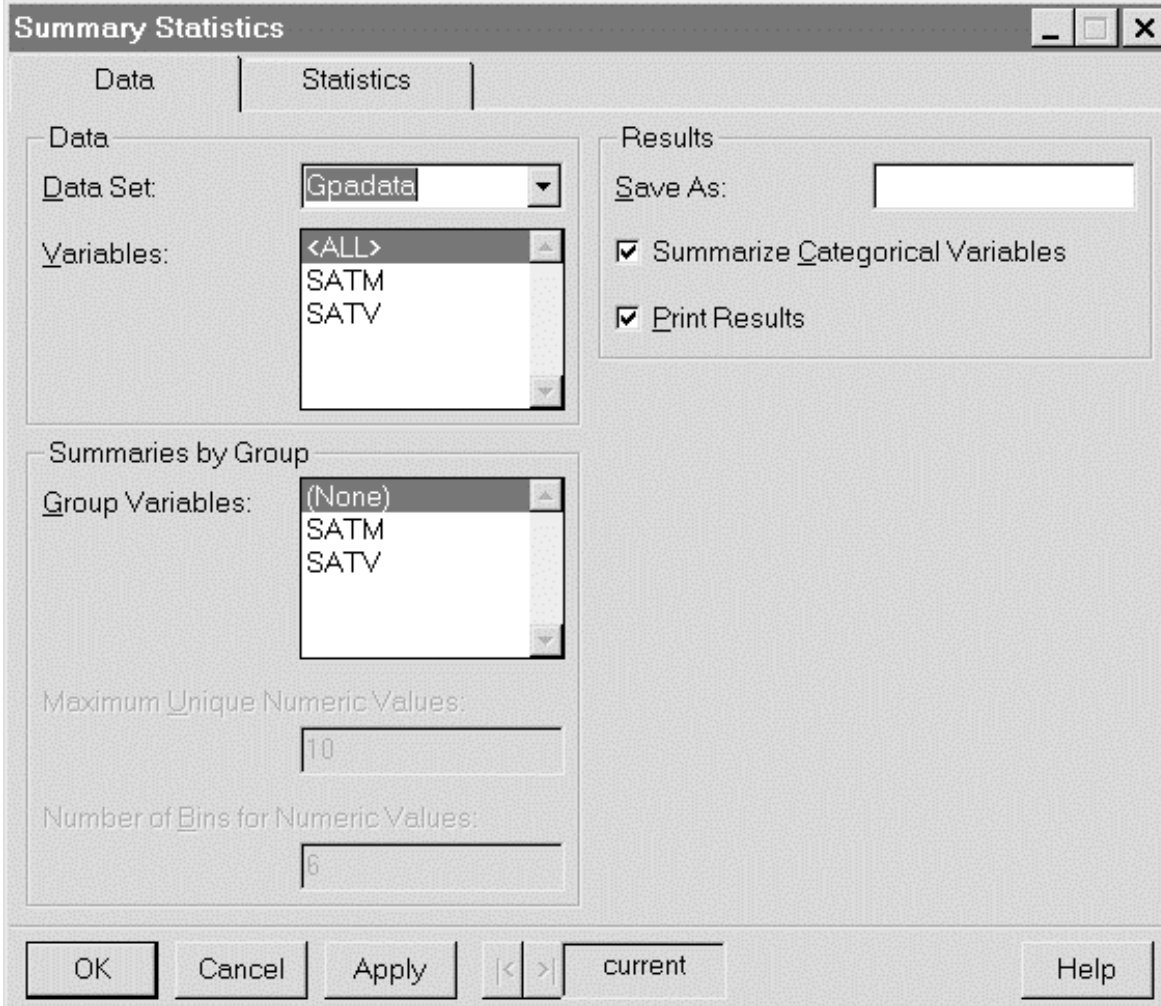
	1	2
	SATM	SATV
1	670.00	600.00
2	630.00	700.00
3	610.00	390.00
4	570.00	530.00
5	700.00	640.00
6	640.00	530.00
7	630.00	560.00
8	610.00	460.00
9	570.00	570.00
10	550.00	500.00
11	670.00	600.00
12	540.00	580.00
13	560.00	690.00
14	630.00	500.00
15	710.00	470.00

Statistics menu options:

- Data Summaries
  - Summary Statistics...
  - Crosstabulations...
  - Correlations...
- Compare Samples
- Power and Sample Size
- Design
- Regression
- ANOVA
- Mixed Effects
- Generalized Least Squares
- Survival
- Tree
- Compare Models...
- Cluster Analysis
- Multivariate
- Quality Control Charts
- Resample
- Smoothing
- Time Series

10. a.

Step 2a



The image shows the 'Summary Statistics' dialog box in SPSS. The 'Data' tab is selected. The 'Data Set' is 'Gpadata'. The 'Variables' list contains '<ALL>', 'SATM', and 'SATV'. The 'Summaries by Group' section has 'Group Variables' set to '(None)'. The 'Maximum Unique Numeric Values' is set to 10, and the 'Number of Bins for Numeric Values' is set to 6. The 'Results' section has 'Save As' set to an empty field, and both 'Summarize Categorical Variables' and 'Print Results' are checked. The bottom buttons are 'OK', 'Cancel', 'Apply', '<', '>', 'current', and 'Help'.

**Summary Statistics**

**Data** | **Statistics**

**Data**

Data Set: Gpadata

Variables: <ALL>  
SATM  
SATV

**Summaries by Group**

Group Variables: (None)  
SATM  
SATV

Maximum Unique Numeric Values: 10

Number of Bins for Numeric Values: 6

**Results**

Save As:

☒ Summarize Categorical Variables

☒ Print Results

OK Cancel Apply < > current Help

10. a.

Step 2b

**Summary Statistics**

Data | Statistics

Mean

- ☒ Mean
- ☐ Std. Error of Mean
- ☐ Conf. Limits for Mean

Conf. Level: 0.95

Quantiles

- ☒ Minimum
- ☒ First Quartile
- ☒ Median
- ☒ Third Quartile
- ☒ Maximum

Scale

- ☐ Variance
- ☒ Std. Deviation

Shape

- ☐ Skewness
- ☐ Kurtosis

Other Statistics

- ☒ Number of Rows
- ☒ Number of Missing Rows
- ☐ Total Sum

OK Cancel Apply < > current Help

10. a.

**Result**

```
S-PLUS - [Report1]
File Edit View Format Data Statistics Graph Options Window Help
[Icons] Linear 1 [Icons]
*** Summary Statistics for data in: Gpadata ***

      SATM
Min: 300.0000000000
1st Qu.: 540.0000000000
Mean: 595.2857142857
Median: 600.0000000000
3rd Qu.: 650.0000000000
Max: 800.0000000000
Total N: 224.0000000000
NA's : 0.0000000000
Std Dev.: 86.4014437443

      SATV
Min: 285.0000000000
1st Qu.: 440.0000000000
Mean: 504.5491071429
Median: 490.0000000000
3rd Qu.: 570.0000000000
Max: 760.0000000000
Total N: 224.0000000000
NA's : 0.0000000000
Std Dev.: 92.6104590705

Ready Modified NUM
```

10. b.

Step 1

S-PLUS - [Gpdata]

File Edit View Insert Format Data Statistics Graph Options Window Help

near 1 6 7

	1	2
	SATM	SATV
1	670.00	600.00
2	630.00	700.00
3	610.00	390.00
4	570.00	530.00
5	700.00	640.00
6	640.00	530.00
7	630.00	560.00
8	610.00	460.00
9	570.00	570.00
10	550.00	500.00
11	670.00	600.00
12	540.00	580.00
13	560.00	690.00
14	630.00	500.00
15	710.00	470.00

Statistics

- Data Summaries
- Compare Samples
- Power and Sample Size
- Design
- Regression
  - Linear...
  - Robust MM...
  - Robust LTS...
  - Stepwise...
  - Generalized Additive...
  - Local (Loess)...
  - Nonlinear...
  - Generalized Linear...
  - Log-linear (Poisson)...
  - Logistic...
  - Probit...
- ANOVA
- Mixed Effects
- Generalized Least Squares
- Survival
- Tree
- Compare Models...
- Cluster Analysis
- Multivariate
- Quality Control Charts
- Resample
- Smoothing
- Time Series

NUM

10. b.

Step 2a

**Formula**

Variable  
Choose Variables: SATM  
SATV

Transformation

Special Term  
Term Category:   
Option:   
Format:   
Add

Add  
Response Main Effect: (+)  
Interaction: (:) Main+Interact: (\*)  
Quadratic: (x^2) Cubic: (x^3)

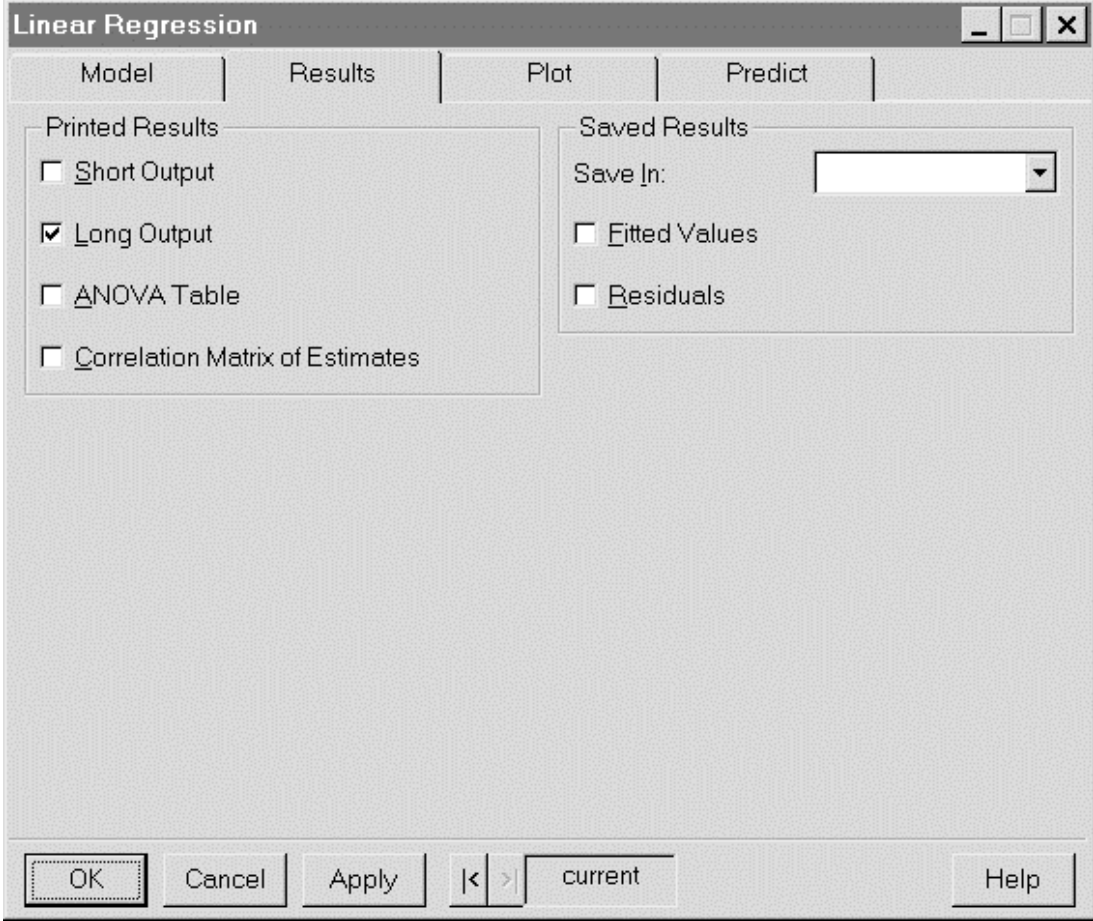
Remove  
☐ Remove Intercept  
Term: SATM  
Remove

Formula: SATV~SATM

OK Cancel Apply |< >| current Help

10. b.

**Step 2**



The image shows a 'Linear Regression' dialog box with the 'Results' tab selected. The 'Printed Results' section on the left contains four options: 'Short Output' (unchecked), 'Long Output' (checked), 'ANOVA Table' (unchecked), and 'Correlation Matrix of Estimates' (unchecked). The 'Saved Results' section on the right has a 'Save In:' dropdown menu and two checkboxes: 'Fitted Values' (unchecked) and 'Residuals' (unchecked). At the bottom, there are buttons for 'OK', 'Cancel', 'Apply', and a navigation section with '<' and '>' arrows and a 'current' label. A 'Help' button is located in the bottom right corner.

Model	Results	Plot	Predict
<b>Printed Results</b>			
<input type="checkbox"/> Short Output			
<input checked="" type="checkbox"/> Long Output			
<input type="checkbox"/> ANOVA Table			
<input type="checkbox"/> Correlation Matrix of Estimates			
<b>Saved Results</b>			
Save In: <span></span>			
<input type="checkbox"/> Fitted Values			
<input type="checkbox"/> Residuals			
OK Cancel Apply  < >  current Help			

10. b.

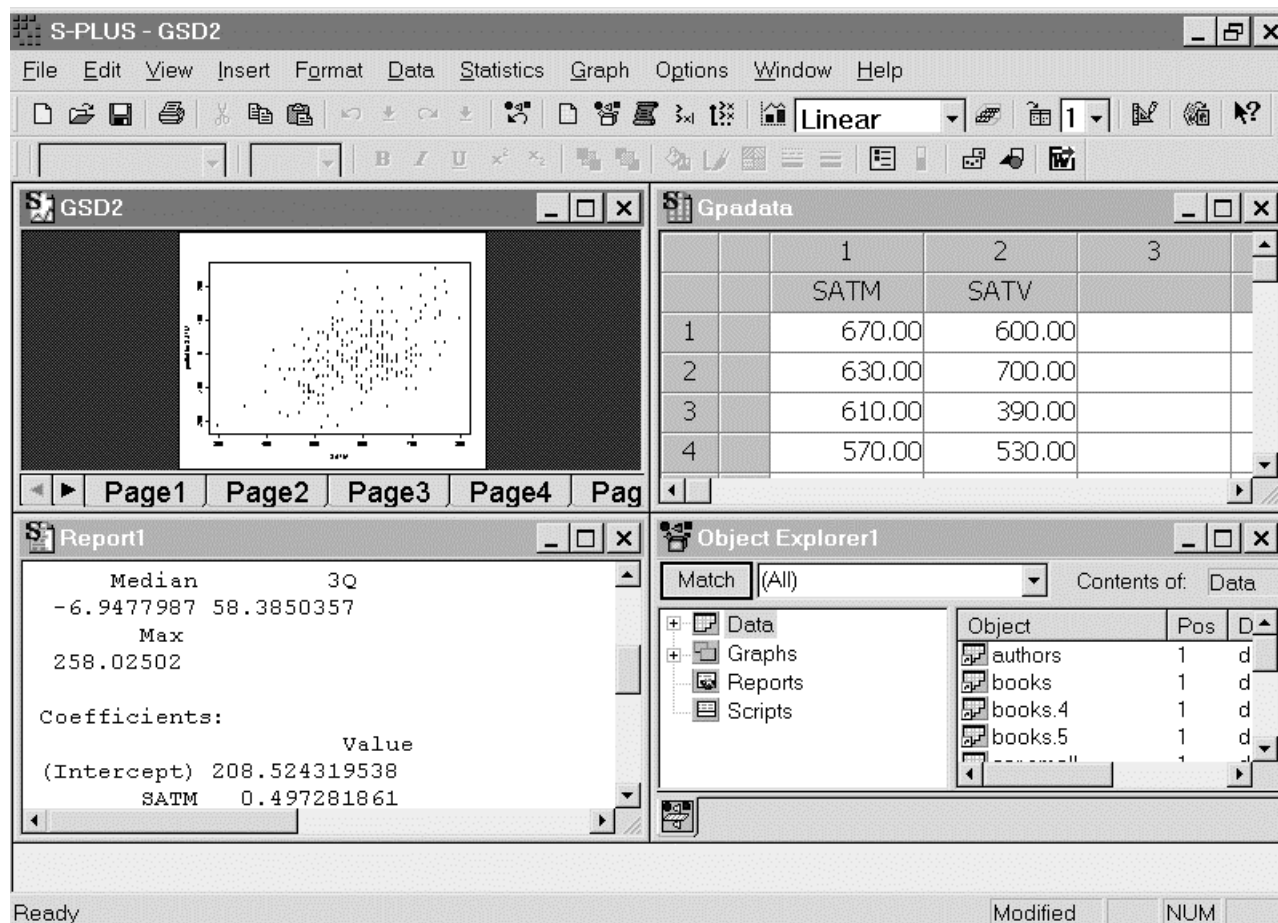
**Step 2**

The screenshot shows the 'Linear Regression' dialog box in Minitab, with the 'Plots' tab selected. The 'Plots' section on the left contains several checked options: 'Residuals vs Fit', 'Sqrt Abs Residuals vs Fit', 'Response vs Fit', 'Residuals Normal QQ', 'Residual-Fit Spread', 'Cook's Distance', and 'Partial Residuals'. The 'Options' section on the right includes 'Include Smooth' (checked), 'Include Rugplot' (unchecked), and a text box for 'Number of Extreme Points To Identify' with the value '3'. Below this, the 'Partial Residual Plot Options' section has 'Include Partial Fit' (unchecked), 'Include Rugplot' (unchecked), and 'Common Y-axis Scale' (checked). At the bottom, there are buttons for 'OK', 'Cancel', 'Apply', navigation arrows, a 'current' label, and a 'Help' button.

Model	Results	Plot	Predict
<b>Plots</b>			
<input checked="" type="checkbox"/> Residuals vs Fit			
<input checked="" type="checkbox"/> Sqrt Abs Residuals vs Fit			
<input checked="" type="checkbox"/> Response vs Fit			
<input checked="" type="checkbox"/> Residuals Normal QQ			
<input checked="" type="checkbox"/> Residual-Fit Spread			
<input checked="" type="checkbox"/> Cook's Distance			
<input checked="" type="checkbox"/> Partial Residuals			
<b>Options</b>			
<input checked="" type="checkbox"/> Include Smooth			
<input type="checkbox"/> Include Rugplot			
Number of Extreme Points To Identify: 3			
<b>Partial Residual Plot Options</b>			
<input type="checkbox"/> Include Partial Fit			
<input type="checkbox"/> Include Rugplot			
<input checked="" type="checkbox"/> Common Y-axis Scale			
OK Cancel Apply  < >  current Help			

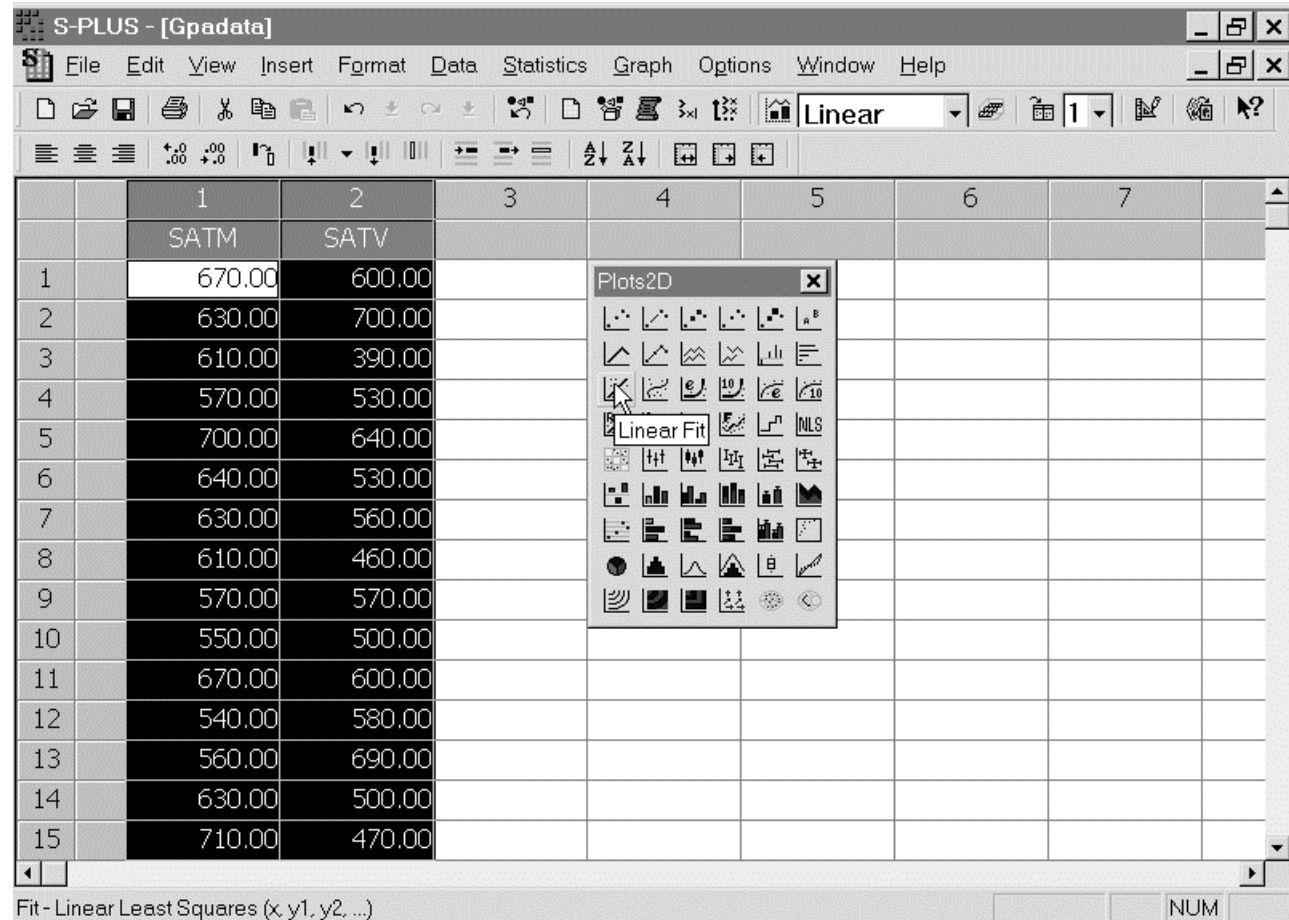
10. b.

**Result**



10. c.

Step 1



10. c.

**Result**

